

VŠB – Technická univerzita Ostrava  
Fakulta elektrotechniky a informatiky  
Katedra informatiky

# **Vícevrstvé sociální sítě**

## **Multilayer Social Networks**

## Zadání diplomové práce

Student: **Bc. Jan Bršťák**

Studijní program: N2647 Informační a komunikační technologie

Studijní obor: 2612T025 Informatika a výpočetní technika

Téma: **Vícevrstvé sociální sítě**  
**Multilayer Social Networks**

Jazyk vypracování: čeština

### Zásady pro vypracování:

Vícevrstvé sociální sítě (jako jsou různá online sociální média), ve kterých různé typy aktérů patří a působí v různých kontextech, jsou atraktivní oblastí výzkumu v oblasti analýzy sociálních sítí. Cílem práce je analýza takovýchto vícevrstevných sítí.

1. Seznamte se s problematikou komplexních sítí, konkrétně sítí vícevrstevných.
2. Seznamte se s možnými způsoby reprezentace vícevrstevných sítí.
3. Naimplementujte aplikaci pro analýzu vícevrstevných sítí.
4. Navrhněte experimenty a nad zvolenými datovými kolekcemi je proveďte. Experimenty vyhodnoťte.

### Seznam doporučené odborné literatury:


- [1] Mark E. Dickison, Matteo Magnani, Luca Rossi: Multilayer Social Networks, 2016, ISBN: 9781107079496, Cambridge University Press.
- [2] Podle pokynů vedoucího diplomové práce.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.


Vedoucí diplomové práce: **RNDr. Eliška Ochodková, Ph.D.**

Datum zadání: 01.09.2017

Datum odevzdání: 30.04.2018

  
doc. Ing. Jan Platoš, Ph.D.  
vedoucí katedry



  
prof. Ing. Pavel Brandštetter, CSc.  
děkan fakulty

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 2018

Beřtal

Rád bych na tomto místě poděkoval vedoucí práce RNDr. Elišce Ochodkové, Ph.D., za odborné rady, vstřícnost, poskytnuté studijní materiály a konzultace.

## **Abstrakt**

Tato diplomová práce popisuje vícevrstvé sociální sítě. Práce podrobně popisuje vlastnosti vícevrstvých sociálních sítí a přístupy, které je možné při analýze těchto sítí zvolit. Pro analýzu vícevrstvých sítí je navržena a implementována aplikace, která umožňuje vícevrstvé sítě analyzovat pomocí několika možných přístupů. Tato aplikace je následně využita při analýze pěti vícevrstvých sítí. Nad těmito sítěmi je provedeno několik experimentů, které se soustředí zejména na porovnání vlastností jednotlivých vrstev a výsledků jednotlivých přístupů k analýze vícevrstvých sítí.

**Klíčová slova:** sítě, vlastnosti sítí, sociální sítě, vícevrstvé sociální sítě, náhodná procházka

## **Abstract**

This master's thesis describes multilayer social networks. Thesis describes properties of multilayer social networks and possible approaches for analysis of these networks in detail. For analysis of multilayer networks is designed and implemented application. This application is able to analyse these networks via several possible approaches. Five multilayer social networks are analysed via this application. Several experiments are performed on these networks. These experiments focus on comparison of networks layers and comparison of results of different approaches that we used for analysis of these multilayer networks.

**Key Words:** networks, network properties, multilayer social networks, random walk

# Obsah

<b>Seznam použitých zkratek a symbolů</b>	<b>8</b>
<b>Seznam obrázků</b>	<b>9</b>
<b>Seznam tabulek</b>	<b>10</b>
<b>1 Úvod</b>	<b>12</b>
1.1 Úvodní příklad . . . . .	13
<b>2 Sítě</b>	<b>15</b>
2.1 Základy teorie grafů . . . . .	16
2.2 Reprezentace grafů . . . . .	18
2.3 Vlastnosti sítí . . . . .	20
<b>3 Vícevrstvé sítě</b>	<b>25</b>
3.1 Model vícevrstvé sociální sítě . . . . .	25
3.2 Další modely vícevrstvých sítí . . . . .	25
<b>4 Vlastnosti vícevrstvých sociálních sítí</b>	<b>29</b>
4.1 Čtyři hlavní přístupy . . . . .	29
4.2 Míry aktérů . . . . .	30
4.3 Flattening . . . . .	39
<b>5 Aplikace</b>	<b>41</b>
5.1 Specifikace . . . . .	41
5.2 Návrh . . . . .	42
5.3 Implementace . . . . .	43
<b>6 Experimenty</b>	<b>49</b>
6.1 Popis datových sad . . . . .	49
6.2 Porovnání jednotlivých vrstev . . . . .	50
6.3 Zjištění role aktéra . . . . .	57
6.4 Porovnání centralit založených na náhodné procházce . . . . .	58
<b>7 Závěr</b>	<b>68</b>
<b>Literatura</b>	<b>69</b>
<b>Přílohy</b>	<b>70</b>



## Seznam použitých zkratek a symbolů

GUI	– Graphical User Interface
XAML	– Extensible Application Markup Language
WPF	– Windows Presentation Foundation
MVVM	– Model-View-ViewModel
API	– Application Programming Interface
RW	– Random Walk



## Seznam obrázků

1	Vícevrstvá sociální síť se čtyřmi vrstvami a osmi aktéry [1] . . . . .	14
2	Příklady grafů . . . . .	17
3	Příklad podgrafu a úplného grafu . . . . .	17
4	Neorientovaný graf . . . . .	18
5	Shlukovací koeficienty . . . . .	21
6	Distribuce stupňů vrcholů . . . . .	22
7	Stupňová centralita vrcholu $A$ . . . . .	23
8	Centralita blízkosti a mezilehlosti . . . . .	24
9	Multiplex vazba . . . . .	26
10	Two-mode síť . . . . .	27
11	Schéma heterogenní informační sítě pro sociální síť Twitter . . . . .	27
12	Temporální síť (sekvence kontaktů) . . . . .	27
13	(a) Náhodná procházka se čtyřmi kroky, kde aktér Serena změní vrstvu LinkedIn na Facebook, jelikož je přítomna na obou z nich, (b) Dvě alternativní cesty $p_1$ a $p_2$ od Marka k Luce [1] . . . . .	30
14	Tři různé cesty mezi Cici a Lucou; (a) a (b) mají stejný počet kroků, ale jiné délky na jiných vrstvách, (c) je delší než (b) [1]. . . . .	35
15	Flattened síť z obrázku 1 . . . . .	39
16	Hlavní okno aplikace . . . . .	44
17	Okno pro porovnání vrstev . . . . .	45
18	Diagram tříd hlavní části aplikace . . . . .	46
19	Kumulativní relativní doplňkové distribuce vrstev vícevrstvé sítě Florentine Families	51
20	Kumulativní relativní doplňkové distribuce všech vrstev vícevrstvé sítě Sampson's Monastery . . . . .	52
21	Kumulativní relativní doplňkové distribuce všech vrstev vícevrstvé sítě Bank Wiring	54
22	Kumulativní relativní doplňkové distribuce všech vrstev vícevrstvé sítě Tailor Shop	55
23	Kumulativní relativní doplňkové distribuce všech vrstev vícevrstvé sítě AUCS . .	56
24	Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě Florentine Families . . . . .	60
25	Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě Sampson's Monastery . . . . .	62
26	Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě Bank Wiring . . . . .	63
27	Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě Tailor Shop . . . . .	65
28	Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě AUCS . . . . .	67

## Seznam tabulek

1	Degree centrality aktérů na specifických vrstvách dle obrázku 1 . . . . .	31
2	Nejkratší vícevrstvé cesty mezi Stine a Markem . . . . .	36
3	Relevance aktérů z obrázku 1 v rámci jednotlivých vrstev [1] . . . . .	38
4	Dataset Florentine Families . . . . .	49
5	Dataset Sampson's Monastery . . . . .	49
6	Dataset Bank Wiring . . . . .	49
7	Dataset Tailor Shop . . . . .	50
8	Dataset AUCS . . . . .	50
9	Základní vlastnosti vrstev vícevrstvé sítě Florentine Families . . . . .	51
10	Vrstvy datasetu Sampson's Monastery . . . . .	53
11	Základní vlastnosti vrstev vícevrstvé sítě Bank Wiring . . . . .	54
12	Základní vlastnosti vrstev vícevrstvé sítě Tailor Shop . . . . .	55
13	Základní vlastnosti vrstev vícevrstvé sítě AUCS . . . . .	56
14	Top 5 aktérů vícevrstvé sítě Florentine Families seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality . . . . .	59
15	Top 5 aktérů vícevrstvé sítě Florentine Families seřazených podle (a) random walk closeness centrality, (b) closeness centrality . . . . .	59
16	Top 5 aktérů vícevrstvé sítě Florentine Families seřazených podle (a) occupation centrality, (b) degree centrality . . . . .	60
17	Top 5 aktérů vícevrstvé sítě Sampson's Monastery seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality . . . . .	61
18	Top 5 aktérů vícevrstvé sítě Sampson's Monastery seřazených podle (a) random walk closeness centrality, (b) closeness centrality . . . . .	61
19	Top 5 aktérů vícevrstvé sítě Sampson's Monastery seřazených podle (a) occu- pation centrality, (b) degree centrality . . . . .	61
20	Top 5 aktérů vícevrstvé sítě Bank Wiring seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality . . . . .	62
21	Top 5 aktérů vícevrstvé sítě Bank Wiring seřazených podle (a) random walk closeness centrality, (b) closeness centrality . . . . .	63
22	Top 5 aktérů vícevrstvé sítě Bank Wiring seřazených podle (a) occupation cent- rality, (b) degree centrality . . . . .	63
23	Top 5 aktérů vícevrstvé sítě Tailor Shop seřazených podle (a) random walk be- tweenness centrality, (b) betweenness centrality . . . . .	64
24	Top 5 aktérů vícevrstvé sítě Tailor Shop seřazených podle (a) random walk clo- seness centrality, (b) closeness centrality . . . . .	64
25	Top 5 aktérů vícevrstvé sítě Tailor Shop seřazených podle (a) occupation centra- lity, (b) degree centrality . . . . .	65

26	Top 5 aktérů vícevrstvé sítě AUCS seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality . . . . .	66
27	Top 5 aktérů vícevrstvé sítě AUCS seřazených podle (a) random walk closeness centrality, (b) closeness centrality . . . . .	66
28	Top 5 aktérů vícevrstvé sítě AUCS seřazených podle (a) occupation centrality, (b) degree centrality . . . . .	66

# 1 Úvod

Téměř každý z nás se pohybuje v mnoha vzájemně propojených sociálních sítích. Tyto sítě nevznikají samovolně, ale jsou vytvářeny našimi sociálními interakcemi a také aktivitami, které tyto sítě propojují a dohromady pak tvoří větší celek, který budeme nazývat vícevrstvá sociální síť. Na online sociálních sítích lidé neustále sdílí fotografie, videa nebo různé informace, které obvykle pocházejí z jiných sítí. Příkladem může být Instagram, který výhradně slouží ke sdílení fotografií, jediné informace, které vznikají na samotném Instagramu jsou komentáře u těchto fotografií. Lidé se neustále, aniž by si to uvědomovali, pohybují v rámci několika sítí, ať už se jedná o sítě offline nebo online. Kdykoli se na sociální síti podělíme o nějakou historku, která se nám zrovna přihodila, přenášíme informaci z jedné sítě do jiné a my jsme tím prostředníkem, který přenos informací mezi těmito sítěmi umožňuje. Bez nás by tato informace na této síti nejspíš nikdy k dispozici nebyla. Některé informace naopak udržujeme jen na určitých sítích, abychom ochránili naše soukromí, udrželi si reputaci, kterou si chceme zachovat před lidmi, kteří nám třeba nejsou tak blízcí, abychom s nimi tuto informaci sdíleli. Fakt, že jsme bez větších obtíží schopni pohybu v rámci několika sítí neznamená, že se jedná o triviální činnost a měli bychom ji ignorovat, právě naopak [1].

Existence několika různých druhů vztahů je známá již dlouhou dobu, lidé si toho byli vědomi již při analýze jednovrstvých sociálních sítí, která byla zaměřena zejména na jeden typ aktérů a vztahů. Ačkoli je tento přístup, kde uvažujeme pouze jeden typ aktéra a vztahu, k analýze sociálních sítí i nadále velmi mocným nástrojem, v posledních letech bylo stále více zřejmé, že tento přístup již není dostačující.

Pohlížet na sociální sítě jako na sítě, ve kterých existuje pouze jeden typ vztahu určitě není správné. Když se podíváme na sociální síť Facebook, i zde existuje několik různých typů vztahů. Uživatelé mohou být označeni za rodinné příslušníky, přátele, nebo mohou pouze sledovat jiné uživatele, popřípadě stránky. Pokud bychom všechny tyto vztahy považovali za přátelství, tak všechny ostatní, které se zde ve skutečnosti nacházejí bychom znehodnotili a vztahy, které jsou tvořeny několika různými typy vztahů, bychom pak úplně odstranili. Tohle na první pohled neškodné zjednodušení může změnit topologii celé sítě, nebo způsob jakým se na vztahy v této síti díváme. Výsledná analýza pak může být nepřesná, v nejhorším případě může dojít i k úplné ztrátě některých výsledků analýzy [1].

Vícevrstvé sociální sítě poskytují zajímavý pohled na jejich aktéry. Například na sociální síti Twitter jsou aktéři reprezentováni pomocí Twitter účtů. Každý účet má počet tweetů, počet sledovaných účtů (following) a počet sledujících účtů (followers). Pokud bychom se rozhodli, že důležitost účtu stanovíme podle počtu sledujících účtů, tak v rámci samotné sítě Twitter to určitě bude v pořádku. Kdybychom ovšem vytvořili vícevrstvou sociální síť, která by obsahovala výše uvedenou sociální síť Twitter a několik dalších sítí, ve kterých je důležitost aktérů vždy stanovena jiným způsobem, tak stanovit důležitost aktérů jen na základě počtu sledujících účtů určitě správně nebude. V každé z těchto sítí je důležitost aktérů stanovena jiným způsobem. Aktéři,

kteří na Twitteru nejsou vůbec důležití, mohou být v rámci některé z dalších sítí naopak velice důležitými. Pokud bychom se na aktéry podívali z trochu odlišné perspektivy a do důležitosti aktérů zahrnuli i další sítě, tak bychom se určitě setkali s takovými aktéry, kteří v rámci Twitteru důležití nebyli, ale v rámci vícevrstvé sociální sítě se důležitými stali [1]. Z tohoto příkladu určitě jasně vidíme, jak zavádějící může být dívat se na sociální sítě pouze z pohledu jedné vrstvy. Kdy pouhým rozšířením naší perspektivy o jednu nebo více dalších vrstev, se může původně nezájímavý aktér stát velmi zajímavým.

Tato nová perspektiva pohledu na analýzu sociálních sítí otevírá nové možnosti, které při analýze jednovrstvých sítí jednoduše nejsou možné. Neměli bychom na ni nahlížet jako na něco, co nahradí analýzu jednovrstvých sítí, ale spíše, jako rozšíření, které je možné na sítě dále aplikovat a získat tak něco nového [1]. Analýza jednovrstvých sítí se velmi osvědčila a určitě bude i nadále velmi využívána, na druhou stranu analýza vícevrstvých sociálních sítí je velmi mladá disciplína a jen čas ukáže, co nám přinese.

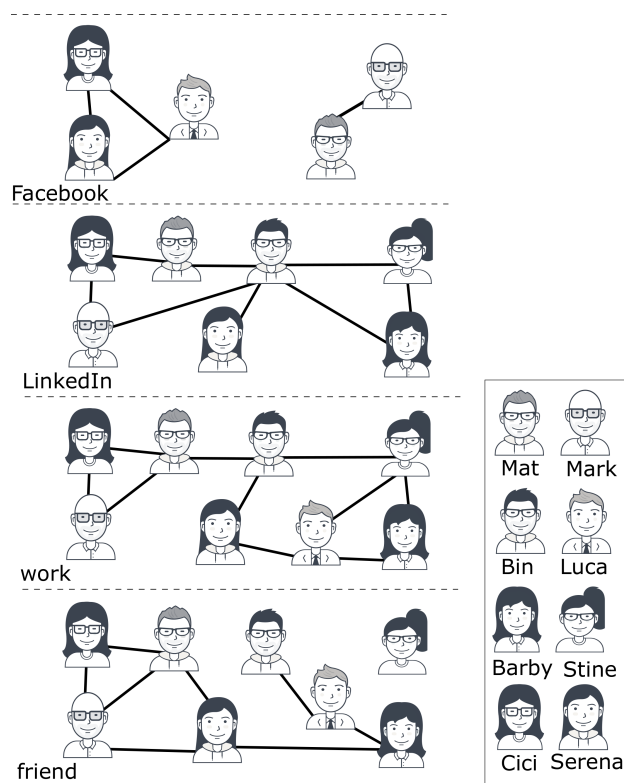
Hlavním zaměřením této práce jsou vícevrstvé sociální sítě, které poskytují nový a velice zajímavý pohled na sociální sítě, především pak jejich aktéry. Cílem práce je tyto sítě přehledně popsat a implementovat některé metody pro jejich analýzu. Následně provést experimenty, ve kterých budou zkoumány vlastnosti aktérů, jednotlivých vrstev a samotné vícevrstvé sítě.

Jelikož vícevrstvé sítě mohou být matoucí, tak je hned v úvodu představena jednoduchá vícevrstvá síť, pomocí které budou napříč několika kapitolami představeny jednotlivé metody, které slouží pro analýzu vícevrstvých sítí. První část práce je pak zaměřena na obecný popis sítí a jejich vlastností. Jsou zde také popsány základy teorie grafů a způsoby reprezentace grafů. Druhá část se zabývá samotnými vícevrstvými sítěmi. Je zde popsáno několik modelů vícevrstvých sítí včetně modelu vícevrstvé sociální sítě. A také jednotlivé vlastnosti vícevrstvých sociálních sítí. Třetí část popisuje implementační část práce, výsledkem které je nástroj pro analýzu vícevrstvých sociálních sítí.

Poslední část je experimentální a obsahuje popis celkem 5 vícevrstvých sítí, nad kterými byly provedeny experimenty. Provedené experimenty se soustředí na porovnání možných přístupů při analýze vícevrstvých sítí. U jednotlivých přístupů jsou porovnávány distribuce centralit jednotlivých vrstev a také konkrétní aktéři a jejich výsledné hodnoty centralit při daném přístupu analýzy.

## 1.1 Úvodní příklad

Tento jednoduchý příklad nám stručně představí model **vícevrstvé sociální sítě** a nové možnosti analýzy, které tento model nabízí. V kapitole 3 se blíže seznámíme s modelem vícevrstvé sociální sítě. Prozatím si takovou síť můžeme představit jako síť obsahující vrcholy a hrany, které jsou uspořádány do několika vrstev, kde každá vrstva reprezentuje jiný typ vrcholu nebo hrany, jiný sociální kontext, jinou komunitu, jinou online sociální síť, atp [1].



Obrázek 1: Vícevrstvá sociální síť se čtyřmi vrstvami a osmi aktéry [1]

Vícevrstvá sociální síť na obrázku 1 je složena ze dvou offline vrstev (work a friend) a dvou online (LinkedIn a Facebook). Všechny tyto vrstvy jsou pak propojeny společnými aktéry, kteří umožňují šíření informací napříč celou vícevrstvou sítí [1].

Pokud se na vícevrstvou síť na obrázku 1 podíváme pozorněji, můžeme si všimnout některých zajímavých vzorů, vznikajících závislostmi mezi těmito čtyřmi vrstvami. Můžeme si všimnout, že aktéři spolu ve spojení nebývají pouze na jedné z vrstev, ale obvykle bývají propojení i na ostatních vrstvách (Cici a Mat). Z tohoto můžeme usoudit, že toto spojení je silnější, než kdyby bylo pouze na jedné vrstvě.

Zajímavý je také aktér Stine, který není ve spojení s nikým na vrstvě friend, i když na vrstvě work má tři spojení. Stine nejspíš nechce sdílet více osobní informace z vrstvy friend. Nechce míchat svůj osobní život s pracovním. Pokud bychom hledali aktéra, který je schopen rozšířit informaci do celé vícevrstvé sítě, tak by to Stine nejspíš nebyla, lepší volbou by mohla být Cici nebo Serena, které jsou na každé ze čtyř vrstev.

Na závěr se můžeme podívat na vzdálenost mezi Cici a Stine. Zajímavé je, že bez nutnosti měnit vrstvy, se mohou vzájemně kontaktovat pouze na vrstvách LinkedIn a work. Cici se ovšem vyskytuje na více vrstvách a nemusí nutně použít přímo LinkedIn nebo work, aby Stine kontaktovala. Může použít Facebook, kde kontaktuje Lucu a ten pak na LinkedInu kontaktuje Stine. Existence těchto vícedimenzionálních vztahů (Cici je přítelem Lucy na Facebooku, který pracuje se Stine) je jedním ze základních nových prostředků v oblasti vícevrstvých sociálních sítí [1].

## 2 Sítě

Sítě můžeme rozdělit na technologické, sociální, informační a biologické. Některé sítě mohou mít aspekty několika z těchto tříd, pak už je jen na nás, jak se tyto sítě rozhodneme klasifikovat. Sítě, které patří do stejné kategorie, jsou pak obvykle analyzovány podobným způsobem [2].

Zástupcem technologických sítí je například Internet. Internet je obrovskou sítí, ve které jsou vrcholy tvořeny počítači nebo síťovými prvky a hrany pak reprezentují fyzické spojení mezi těmito vrcholy. Toto spojení může být realizováno například pomocí optického vlákna [2].

V sociálních sítích jsou vrcholy obvykle tvořeny lidmi nebo skupinou lidí. Z tohoto důvodu se také vrcholy v sociálních sítích nazývají aktéři, hrany se pak nazývají vazby [2]. Jako představitele sociální sítě si uvedeme Twitter. Aktéři jsou zde reprezentováni pomocí Twitter účtů. Vazba mezi dvěma aktéry bude existovat tehdy, pokud alespoň jeden z aktérů sleduje toho druhého, jedná se tedy o orientované vazby, o kterých se zmíníme později.

Informační sítě jsou tvořeny daty, které jsou nějakým způsobem vzájemně propojeny. Všechny informační sítě, o kterých prozatím víme, byly vytvořené lidmi, nejznámější informační sítí je pak bezesporu World Wide Web. V této síti jsou vrcholy tvořeny webovými stránkami, a hrany tvoří hypertextové odkazy, které nám umožňují navigaci napříč těmito stránkami. Jelikož tyto odkazy odkazují z jedné stránky na jinou, tak jsou tyto hrany orientované. Mezi informační sítě patří také citační sítě. Tyto sítě jsou mnohem starší, než World Wide Web, ale velmi podobné. Vrcholy zde mohou být tvořeny například články a hrany jsou pak citované zdroje [2]. Jedná se pak opět o orientované hrany.

Biologické sítě jsou pak využívány například v neurovědě pro reprezentaci propojení mozkových buněk [2].

Všechny z výše uvedených sítí mohou mít velmi zajímavé vlastnosti, které nám mohou povědět více o struktuře dané sítě. Při analýze sítí se velmi často setkáme s distribucí stupňů vrcholů, shlukovacím koeficientem, délkou cest mezi vrcholy a dalšími zajímavými vlastnostmi, které nám o sítích mohou hodně prozradit, pokud víme, jak je interpretovat. S těmito a také řadou jiných vlastností se blíže seznámíme v následujících kapitolách.

**Náhodný graf** je jednoduše řečeno graf, který byl vytvořen pomocí nějakého náhodného procesu. Náhodné grafy byly poprvé definovány a popsány maďarskými matematiky Paulem Erdősem a Alfrédem Rényim. Dnes známe tzv. Erdős–Rényi model náhodných grafů. V tomto modelu je pevně dán počet vrcholů  $n$  a mezi tyto vrcholy je s uniformní pravděpodobností vloženo  $m$  hran. Tuto variantu označujeme jako  $G(n, m)$ . Varianta  $G(n, p)$ , pak má, stejně jako předchozí, pevně dán počet vrcholů  $n$  a  $p$  pak udává pravděpodobnost vložení hrany mezi libovolnou dvojici vrcholů [14].

Rozsáhlé sítě, které mají reálný původ, označujeme jako **komplexní sítě**, ty se od náhodných grafů (sítí) odlišují především svou komplexní, nepravidelnou a dynamicky se v čase rozvíjející strukturou. Kvůli těmto vlastnostem může být velmi obtížné tyto sítě analyzovat. Jako zástupce komplexních sítí si můžeme představit například World Wide Web, který se od svého vzniku

změnil téměř k nepoznání. Komplexní sítě a především jejich topologie se staly velmi populárním předmětem výzkumu, jelikož samotná struktura může ovlivňovat procesy v těchto sítích. Například u sociálních sítí dokáže topologie výrazně ovlivnit rychlost šíření informace a můžeme z ní vyčíst, kteří aktéři v těchto procesech hrají důležitou roli [3].

Při analýze sítí se téměř vždy setkáme s grafy, pomocí kterých lze velice snadno a efektivně sítě zobrazit. Podívejme se tedy na základní pojmy teorie grafů.

## 2.1 Základy teorie grafů

Za zakladatele teorie grafů je považován Leonhard Euler, který v roce 1736 dokázal, že projít každý ze sedmi mostů v Královci pouze jednou a vrátit se do původního místa není možné a zavedl tak pojem eulerovský tah [4]. Od onoho roku už uplynula dlouhá doba a v teorii grafů bylo zavedeno mnoho nových pojmů.

### 2.1.1 Graf

Jedním z pojmů, se kterým se v teorii grafů zaručeně setkáme, je jednoduchý graf. Pomocí grafu lze snadno graficky vyjádřit vztahy mezi objekty. Tyto objekty jsou v grafech reprezentovány pomocí vrcholů, vztahy mezi vrcholy jsou pak reprezentovány pomocí hran.

**Definice 2.1** *Jednoduchý graf  $G$  je uspořádaná dvojice  $(V, E)$ , kde  $V$  je neprázdná množina vrcholů a  $E$  je množina dvouprvkových podmnožin množiny  $V$  [5].*

Hrany grafu (sítě) mohou mít také vlastnosti, které určují jeho charakter. Grafy tak lze dále dělit:

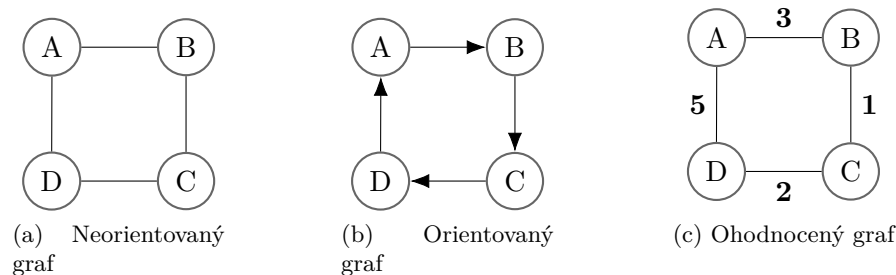
- Podle orientace hran
  - Neorientovaný graf
  - Orientovaný graf
- Podle ohodnocení hran
  - Ohodnocený graf
  - Neohodnocený graf

**Definice 2.2** *Neorientovaný graf je definován jako uspořádaná dvojice  $G = (V, E)$ , kde  $V$  je neprázdná množina vrcholů a  $E$  je množina neuspořádaných dvojic  $u, v \in V$  [5].*

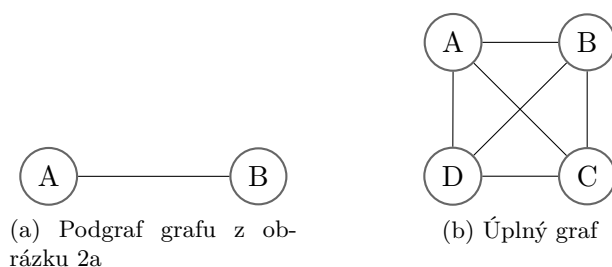
Na obrázku 2a pak můžeme vidět neorientovaný graf se čtyřmi vrcholy a čtyřmi hranami.

**Definice 2.3** *Orientovaný graf je definován jako uspořádaná dvojice  $G = (V, E)$ , kde  $V$  je neprázdná množina vrcholů a  $E$  je množina uspořádaných dvojic  $(u, v) \in V$  [5].*





Obrázek 2: Příklady grafů



Obrázek 3: Příklad podgrafu a úplného grafu

Příklad orientovaného grafu se čtyřmi vrcholy a čtyřmi hranami je pak uveden na obrázku 2b.

**Definice 2.4** Ohodnocení grafu  $G$  je funkce  $w : E(G) \rightarrow \mathbb{R}$ , která každé hraně  $e \in E(G)$  přiřadí reálné číslo  $w(e)$ , kterému říkáme váha hrany. Ohodnocený graf je graf  $G$  spolu s ohodnocením hran reálnými čísly [5].

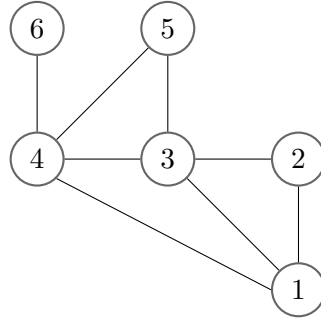
Ohodnocený graf je tedy graf, který hranám přiřazuje nenulová čísla. Na obrázku 2c je pak znázorněn ohodnocený graf. Neohodnocený graf pak naopak hranám nenulová čísla nepřisazuje. Graf na obrázku 2a je tedy také neohodnoceným grafem.

**Definice 2.5** Podgraf grafu  $G$  je graf  $G'$ , který vznikne odebráním některých vrcholů a hran z původního grafu  $G$ . Nejčastěji se můžeme setkat s označením  $G' = (V', E')$ , kde  $V' \subseteq V$  a  $E' \subseteq E$  [5].

**Definice 2.6** Úplný graf je takový neorientovaný graf, ve kterém jsou každé dva vrcholy spojené hranou. Obsahuje tak při daném počtu vrcholů maximální počet hran [5].

**Definice 2.7** Stupeň vrcholu  $v$  je počet hran, se kterými je vrchol  $v$  incidentní, a značí se  $\deg(v)$  [5].

**Definice 2.8** Vrcholy  $u, v \in V$  nazýváme sousední, jestliže mezi nimi existuje hrana, která tyto dva vrcholy spojuje. Vrcholy  $u, v$  jsou tak incidentní s hranou  $\{u, v\}$  [5].



Obrázek 4: Neorientovaný graf

**Definice 2.9** *Sled v grafu  $G$  je taková posloupnost vrcholů a hran  $(v_0, e_1, v_1, e_2, v_2, \dots, e_n, v_n)$ , že hrana  $e_i$  má koncové vrcholy  $v_{i-1}$  a  $v_i$  pro všechna  $i = 1, 2, \dots, n$  [5].*

Sled je tedy posloupnost vrcholů, kde pro vrchol  $v_i$  a vrchol následující  $v_{i+1}$  existuje hrana.

**Definice 2.10** *Tah je sled, ve kterém se žádná hrana neopakuje [5].*

**Definice 2.11** *Cesta v grafu je sled, ve kterém se neopakují vrcholy [5].*

## 2.2 Reprezentace grafů

Existuje několik způsobů jakými lze grafy reprezentovat, každý z těchto způsobů má své výhody i nevýhody. Podívejme se teď na tři základní způsoby, kterými lze neorientované grafy reprezentovat.

### 2.2.1 Seznam hran

Jedním z nejjednodušších způsobů, kterým lze graf reprezentovat, je pomocí seznamu hran. Hranu budeme reprezentovat pomocí dvojice vrcholů, kterou můžeme uložit např. do pole, nebo objektu. Všechny hrany, které jsou součástí grafu pak uložíme do seznamu. Pokud bychom potřebovali reprezentovat i váhy hran, můžeme místo dvojic hran ukládat trojice, kde třetím prvkem bude váha dané hrany [5]. Jelikož se každá hrana skládá ze dvou vrcholů, popřípadě dvou vrcholů a váhy hrany, tak celková paměťová složitost seznamu hran je  $\Theta(E)$ . Graf z obrázku 4 reprezentovaný pomocí seznamu hran (1).

$$[[v_1, v_2], [v_1, v_3], [v_1, v_4], [v_2, v_3], [v_3, v_4], [v_3, v_5], [v_4, v_5], [v_4, v_6]] \quad (1)$$

### 2.2.2 Seznam sousedů

Reprezentace grafu pomocí seznamu sousedů kombinuje matice sousednosti spolu se seznamy hran [5]. Každý vrchol  $i$  má své sousedy uložené např. v poli, nebo seznamu. Graf z obrázku 4 reprezentovaný pomocí seznamu sousedů (2).

$$\begin{aligned}
 v_1 &= [v_2, v_3, v_4] \\
 v_2 &= [v_1, v_3] \\
 v_3 &= [v_1, v_2, v_4, v_5] \\
 v_4 &= [v_1, v_3, v_5, v_6] \\
 v_5 &= [v_3, v_4] \\
 v_6 &= [v_4]
 \end{aligned} \tag{2}$$

Každý seznam sousedů vrcholu jsme schopni získat v konstantním čase  $O(1)$ . Ke zjištění existence hrany  $(i, j)$  se podíváme na seznam sousedů vrcholu  $i$  a zjistíme jestli se zde nachází vrchol  $j$ . Složitost nalezení hrany je tedy  $\Theta(d)$ , kde  $d$  je stupeň daného vrcholu. Stupeň vrcholu může být nejvýše  $|V| - 1$  a nejméně pak 0. Co se týče paměťové náročnosti, tak celkově máme  $|V|$  seznamů sousedů, kde každý z těchto seznamů může obsahovat až  $|V| - 1$  vrcholů, celkem tedy seznam sousedů pro neorientované grafy obsahuje  $2|E|$  prvků. Jelikož se zde každá hrana nachází dvakrát.

### 2.2.3 Matice sousednosti

Pro reprezentaci grafu  $G$  s  $n$  vrcholy pomocí matice sousednosti  $A(G)$  budeme potřebovat čtvercovou matici řádu  $n$ , ve které prvek na řádku  $i$  a sloupci  $j$  bude nabývat hodnoty 1 pouze pokud v grafu existuje hrana  $e(i, j)$ , jinak 0 [5]. Graf z obrázku 4 reprezentovaný pomocí matice sousednosti (3).

$$A(G) = \begin{matrix} & \begin{matrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \end{matrix} \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{matrix} \tag{3}$$

Pokud bychom chtěli ukládat také váhy hran, můžeme místo hodnoty 1 ukládat samotnou hodnotu váhy hrany a rezervovat si speciální hodnotu indikující nepřítomnost hrany. Hlavní výhodou matice sousednosti je zjištění přítomnosti hrany v konstantním čase  $O(1)$ . Jedna z nevýhod je paměťová složitost. Pro uložení i relativně řídkého grafu, tedy grafu, který má relativně

málo hran vůči počtu vrcholů bude  $\Theta(V^2)$ . Další z nevýhod je nalezení sousedů vrcholu. Pokud chceme nalézt všechny sousedy vrcholu  $v_i$  musíme projít všechny vrcholy na řádku  $i$  i kdyby tento vrchol měl pouze jednoho souseda, složitost je  $O(|V|)$ . Přidávání a odebrání vrcholů a hran je zde také velmi náročná operace, jelikož je nutné znovu sestavit matici sousednosti.

## 2.3 Vlastnosti sítí

V předchozí kapitole jsme se primárně zaměřili na grafy a jejich možné reprezentace, jelikož i sítě jsou grafy a se všemi výše zmíněnými pojmy se setkáme i u sítí. V této kapitole se více zaměříme na sítě a hlavně jejich vlastnosti, se kterými se u náhodných grafů nebo sítí nejspíše nesetkáme.

### 2.3.1 Fenomén malého světa

Jednou z velmi zajímavých vlastností, kterou se sítě vyznačují, je jejich relativně malý průměr vůči celkovému počtu vrcholů. Tuto vlastnost nazýváme **fenomén malého světa**, pro průměrnou vzdálenost u těchto sítí pak platí vzorec 4.

$$l \propto \ln N \quad (4)$$

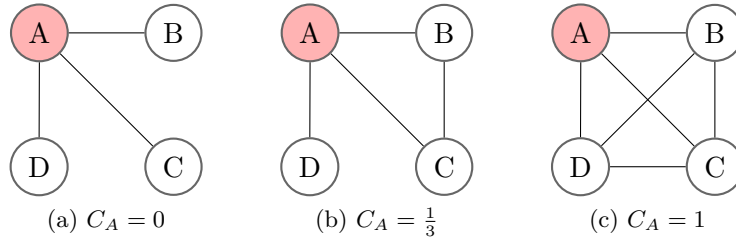
**Definice 2.12** *Vzdálenost dvou vrcholů  $u$  a  $v$  je délka nejkratší cesty mezi  $u$  a  $v$ , značí se pak jako  $d(u, v)$  [5].*

**Definice 2.13** *Průměrná vzdálenost je průměrem nejkratších cest mezi všemi dvojicemi vrcholů sítě. Pokud mezi dvojicí vrcholů  $u$  a  $v$  cesta neexistuje nebo  $u = v$ , pak je jejich vzdálenost rovna nule [15].*

$$l = \frac{1}{N(N-1)} \sum_{u,v \in V} d(u, v) \quad (5)$$

Bylo provedeno mnoho experimentů, které tuto vlastnost potvrzují. Experiment, který vedl k samotnému objevu **fenoménu malého světa** byl proveden v roce 1967. Provedl ho Stanley Milgram z Harvardovy univerzity, který náhodně vybral více než stovku osob, které následně požádal, aby se pokusily doručit zprávu adresovanou jeho příteli pouze prostřednictvím osob, se kterými se osobně znají. Celkově bylo odesláno 160 dopisů a doručeno jich bylo 44. Počet osob nutných k úspěšnému doručení zprávy se pohyboval od 2 do 10, průměrný počet pak byl 6 [6]. Z těchto výsledků pak vyvodil závěr, že pro kontaktování libovolné jiné osoby stačí v průměru pouze šest osob. Tato teorie se pak nazývá „šest stupňů odloučení“ [7].

V roce 2011 byl proveden další ze zajímavých experimentů. V sociální síti Facebook mezi 721 milióny aktivními uživateli byla průměrná vzdálenost stanovena na 3,74, o pár let později v roce 2016 pak byla tato vzdálenost znovu změřena, ovšem už mezi 1,59 miliardou aktivních uživatelů a vzdálenost už byla rovna „pouze“ 3,57. S postupným růstem sítě se uživatelé stávali



Obrázek 5: Shlukovací koeficienty

více propojenými a průměrná vzdálenost se snížila [9]. V citačních sítích jsou od sebe autoři vzdáleni 4 až 6 vazbami. Na Wikipedii je v síti článků průměrná vzdálenost 3,45 [16]. Fenomén malého světa lze tedy považovat za jednu z typických vlastností sítí.

### 2.3.2 Shluky

Shluky jsou vrcholy, které jsou navzájem hustě propojené. Jedná se o další z typických vlastností reálných sítí. Lidé mají tendenci tvořit shluky. Je velice pravděpodobné, že pokud známe Petra a Pavla, tak Petr také zná Pavla. Tedy pokud existuje hrana mezi mnou, Petrem a Pavlem, tak nejspíše také existuje hrana mezi Petrem a Pavlem.

Důležitou mírou, se kterou se s velkou pravděpodobností při shlukové analýze setkáme je **shlukovací koeficient**. Shlukovací koeficient je míra, která počítá hustotu uzavřených cest délky tři, tedy hustotu trojúhelníků. Můžeme se setkat se shlukovacím koeficientem globálním 6 a lokálním 7 [2].

$$C = \frac{3 \times \text{celkový počet trojúhelníků}}{\text{počet trojic propojených dvěma hranami}} \quad (6)$$

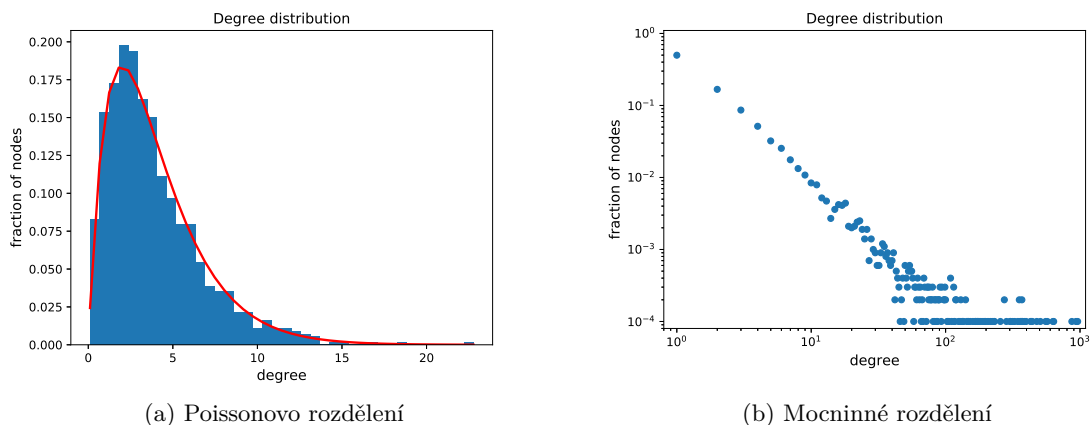
$$C_i = \frac{\text{počet propojených párů sousedů vrcholu } i}{\text{celkový počet párů sousedů vrcholu } i} \quad (7)$$

Alternativní způsob výpočtu shlukovacího koeficientu je pak definován vztahem 8, kde  $L_i$  je počet hran mezi sousedy vrcholu  $i$  a  $k_i$  je celkový počet sousedů vrcholu  $i$ . Tento způsob bývá často upřednostňován především díky nižší výpočetní náročnosti. Za pomoci tohoto vztahu lze snadno získat **průměrný shlukovací koeficient** 9, kde  $N$  je počet vrcholů sítě [3].

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \quad (8)$$

$$C_i = \frac{1}{N} \sum_{i=1}^N C_i \quad (9)$$

I v tomto případě byly provedeny experimenty, které potvrzují, že reálné sítě skutečně mají výrazně vyšší shlukovací koeficient, než náhodné sítě. Například síť spolupráce fyziků, která



Obrázek 6: Distribuce stupňů vrcholů

obsahuje celkem 53000 vrcholů má shlukovací koeficient  $C = 0,45$ . Sít spolupráce herců, která obsahuje 450000 vrcholů má  $C = 0,2$ . Jedná se o síť s výrazně vyšším počtem trojúhelníků, než je běžné u náhodných sítí [2].

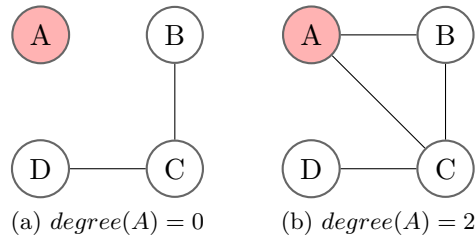
### 2.3.3 Mocninný zákon

Další z vlastností, na kterou se v této kapitole zaměříme je **mocninný zákon**. Na rozdíl od náhodných sítí, kde má většina vrcholů stupeň podobný průměrnému stupni. Jedná se tedy o binomickou distribuci stupňů vrcholů, kterou lze aproximovat pomocí Poissonova rozdělení (viz obrázek 6a). Reálné sítě mají distribuci stupňů podle mocninného zákona (viz obrázek 6b), kde vidíme, že zde opravdu je několik vrcholů s výrazně vyšším stupněm. Tyto vrcholy pak nazýváme **centra**. Sítě, které mají distribuci stupňů vrcholů podle mocninného zákona, se pak nazývají **bezškálové** [17].

Zajímavý experiment provedl Maďarský vědec Albert-László Barabási, který zmapoval celkem 325 000 stránek z domény nd.edu a zjistil, že na některé stránky odkazuje mnohem více stránek, než na ostatní. Sít respektovala distribuci stupňů vrcholů dle mocninného zákona. Následně provedl několik dalších experimentů, a zjistil že i další sítě, včetně sociálních a biologických se také řídí mocninným zákonem [17]. Se svou kolegyní Rékou Albert pak představili Barabási-Albert model, který je založen na preferenčním připojování, kde se nové vrcholy připojují preferenčně k vrcholům s vyšším stupněm [8]. Díky preferenčnímu připojování tento model také zachovává mocninný zákon stupňů vrcholů a umožňuje vznik center.

### 2.3.4 Centrality

V předchozí kapitole jsme trochu nastínili existenci center v reálných sítích. Zatím jsme tyto centra označili jako vrcholy, které se od ostatních liší především svým výrazně vyšším stupněm. Pokud bychom ovšem důležitost vrcholů neurčovali jen na základě jejich stupňů, i jiné vrcholy by



Obrázek 7: Stupňová centralita vrcholu  $A$

se mohly stát důležitými. Podívejme se teď na několik centralit, které se pro měření důležitosti vrcholů běžně používají.

**Stupňová centralita (Degree Centrality)** je základní centralita, která určuje důležitost vrcholu pouze na základě jeho stupně v síti. Pokud by vrcholem byl uživatel sociální sítě Facebook a hrana by byla reprezentována přátelstvím, tak by nejdůležitějším uživatelem byl ten s největším počtem přátel.

**Centralita blízkosti (Closeness Centrality)** měří důležitost vrcholu na základě průměrné hodnoty vzdálenosti od všech ostatních vrcholů v síti. Nicméně při tomto výpočtu důležité vrcholy mají malé hodnoty a naopak vrcholy, které důležité nejsou mají hodnoty vysoké, z tohoto důvodu se tato centralita počítá jako inverzní hodnota tohoto průměru. Vrchol, který je z pohledu této míry důležitý, může mít dobrý přístup k ostatním vrcholům a může rychle rozšířit informaci v síti. Průměrnou vzdálenost vrcholu  $v_i$  od ostatních vrcholů pak lze spočítat pomocí vzorce [2]:

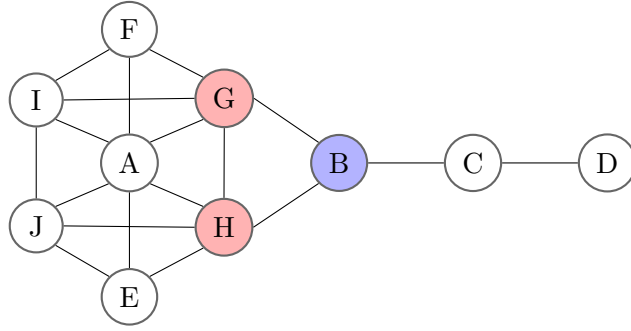
$$l_i = \frac{1}{N} \sum_j d_{ij}, \quad (10)$$

kde  $N$  je počet vrcholů v síti a  $d_{ij}$  je nejkratší cesta mezi vrcholy  $v_i$  a  $v_j$ . Samotnou centralitu pak lze spočítat pomocí vzorce [2]:

$$C_i = \frac{1}{l_i}, \quad (11)$$

kde  $l_i$  je průměrná vzdálenost vrcholu  $i$  ke všem ostatním vrcholům. Na obrázku 8 pak lze snadno určit, že nejdůležitější vrcholy z pohledu centrality blízkosti jsou vrcholy  $G$  a  $H$ , jelikož mají jednoznačně nejblíže ke všem ostatním vrcholům.

**Centralita mezilehlosti (Betweenness Centrality)** k měření důležitosti vrcholu přistupuje odlišným způsobem, než předchozí dvě centrality. Důležitost vrcholu je zde určena podle počtu nejkratších cest mezi dvěma různými vrcholy (počítáno přes všechny možné dvojice vrcholů), na kterých vrchol leží. Vrcholy s vysokou hodnotou betweenness mohou mít velký vliv na pohyb v síti a mohou tedy například kontrolovat, jaké informace se dostanou k jakému vrcholu. Při odstranění vrcholu s vysokou hodnotou betweenness může dojít k výraznému zpomalení šíření informací v síti, jelikož daný vrchol mohl být vrcholem, přes který informace mohly putovat nejrychleji. U předešlých centralit by takový vrchol nemusel být považován za důležitý, avšak



Obrázek 8: Centralita blízkosti a mezilehlosti

jeho odstraněním můžeme celou síť výrazně ovlivnit. Betweenness vrcholu  $v_i$  lze spočítat pomocí vzorce 12, kde  $g_{st}$  je počet všech nejkratších cest mezi vrcholy  $v_s$  a  $v_t$ , a  $n_{st}^i$  je počet nejkratších cest, které vedou přes vrchol  $v_i$ . Na obrázku 8 pak lze stejně, jako v předchozím případě jednoduše identifikovat nejdůležitější vrchol z pohledu betweenness centrality, jelikož největší množství nejkratších cest jednoznačně vede právě přes vrchol  $B$ .

$$B_i = \sum_{st} \frac{n_{st}^i}{g_{st}} \quad (12)$$



### 3 Vícevrstvé sítě

V kapitole 1 jsme si už něco málo o vícevrstvých sítích pověděli, zejména pak o vícevrstvých sociálních sítích. Vícevrstvé sítě byly předmětem výzkumu mnoha oborů. I když každý z těchto oborů měl odlišný přístup k analýze vícevrstvých sítí, tak nakonec musely řešit podobné problémy. Toto vedlo k vzniku několika modelů, kterými lze vícevrstvé sítě reprezentovat, několika přístupů k samotné analýze a mnoha vlastností, které lze u vícevrstvých sítí zkoumat, na které se v následujících kapitolách blíže zaměříme.

#### 3.1 Model vícevrstvé sociální sítě

V následující kapitole se seznámíme s několika modely vícevrstvých sítí. Nicméně, pokud nespecifikujeme, že se jedná o některý z těchto modelů, tak za model vícevrstvé sítě budeme považovat model definovaný následující definicí:

**Definice 3.1** *Vícevrstvá síť je definovaná jako čtveřice  $M = (A, L, V, E)$ , kde  $A$  je množina aktérů,  $L$  je množina vrstev,  $(V, E)$  je graf, kde  $V \subseteq A \times L$  [1].*

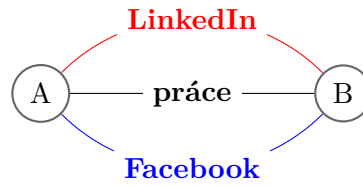
Vícevrstvá síť se tedy skládá z množin aktérů a vrstev, kde dvojice složená aktérem a vrstvou tvoří vrchol. A stejně jako v běžné síti, tak i zde mohou být vrcholy propojeny hranou.

Jelikož vícevrstvá síť obsahuje několik vrstev, je nutné rozlišovat různé typy hran. Hrany, které propojují vrcholy v rámci stejné vrstvy nazýváme **intralayer**. Pokud pak hrany propojují vrcholy v rámci odlišných vrstev, nazýváme je **interlayer**. Na každé z vrstev mohou intralayer hrany reprezentovat jiný druh vazby, například na vrstvě reprezentující práci může hrana představovat spolupráci, zatímco na jiné vrstvě přátelství (sociální síť Facebook).

Tok informací napříč různými vrstvami je pak realizován pomocí interlayer hran. Pro lepší porozumění významu těchto hran si představme vícevrstvou síť, která obsahuje dvě vrstvy reprezentující pracovní a osobní e-mailové adresy. Dva zaměstnanci stejné společnosti mohou komunikovat pomocí pracovních e-mailových adres (reprezentováno pomocí intralayer hrany na pracovní vrstvě), jeden ze zaměstnanců si může přeposílat e-maily z pracovní adresy na osobní (reprezentováno pomocí interlayer hrany z pracovní vrstvy na osobní). V extrémním případě může zaměstnanec text obsažený v pracovním e-mailu jednoduše zkopírovat, vložit ho do svého osobního e-mailu a odtud text odeslat na osobní e-mail druhého zaměstnance. V tomto případě žádná informace neprošla přes interlayer hrany, zaměstnanec pouze zaměnil svou vrstvu. Vícevrstvé sítě nám umožňují modelovat toto komplexní chování aktérů (přechody mezi různými vrstvami), které se odvíjí od jejich záměrů a možností, které jednotlivé vrstvy poskytují [1].

#### 3.2 Další modely vícevrstvých sítí

V této kapitole se blíže seznámíme s několika vybranými modely vícevrstvých sítí.



Obrázek 9: Multiplex vazba

### 3.2.1 Multiplex síť

V multiplex sítích, někdy se také můžeme setkat s názvy multidimenzionální nebo multirelační sítě, jsou vrcholy propojeny pomocí několika typů hran, jak je ukázáno na obrázku 9. Vztah mezi vrcholy  $A$  a  $B$  je zde reprezentován třemi typy hran. Vrcholy  $A$  a  $B$  jsou tedy přátelé na Facebooku (modrá hrana), pracují ve stejné společnosti (černá hrana) a také jsou ve spojení na síti LinkedIn (červená hrana).

Multiplex síť jsou jednou z nejstarších reprezentací vícevrstevných sítí, obvykle jsou pak reprezentovány pomocí několika matic sousednosti 2.2.3, kde každý typ hrany je reprezentován jednou maticí. Už při této reprezentaci si lidé začali uvědomovat, že počet vazeb mezi dvěma vrcholy je přímo úměrný síle propojení těchto vrcholů. Multiplex vazbu lze tedy považovat za silnější, jelikož propojuje vrcholy pomocí několika typů spojení [1]. Pokud jsou dva lidé dobrými přáteli a navíc pracují ve stejné společnosti, s velkou pravděpodobností zůstanou v kontaktu i kdyby jedna z těchto vazeb byla odstraněna.

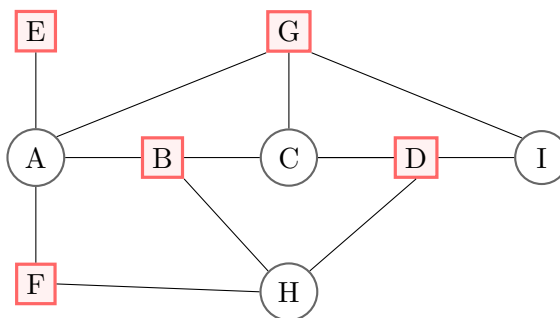
### 3.2.2 Multimode a Multilevel síť

Zatímco multiplex síť využívají různé typy hran, multimode a multilevel síť využívají různé typy vrcholů. U multimode sítí se nejčastěji můžeme setkat s variantou se dvěma typy vrcholů, tyto síť se pak nazývají two-mode síť (můžeme se také setkat s označením bipartitní síť) [1]. Hran v těchto sítích pak vždy vedou pouze mezi odlišnými typy vrcholů, jak můžeme vidět na obrázku 10. Tato síť pak může představovat například společnost a zaměstnance, nebo univerzitu a studenty, kde vrchol  $A$  je společnost/univerzita a ostatní vrcholy jsou zaměstnanci/studenti. Multilevel síť jsou pak rozšířením two-mode sítí. Tyto síť navíc umožňují existenci hran mezi stejnými typy vrcholů [1].

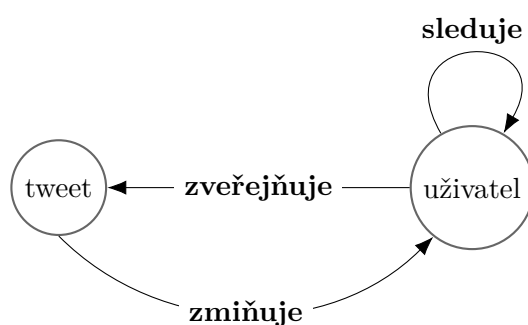
### 3.2.3 Heterogenní informační síť

Informační síť reprezentují abstrakci skutečného světa, zaměřují se na objekty a interakce mezi nimi. Tato úroveň abstrakce slouží pro uchování základních informací o skutečném světě a také poskytuje užitečný nástroj pro nalezení vzorů, které se zde mohou vyskytovat.

Pokud informační síť obsahuje více než jeden druh vrcholu nebo vazby nazýváme ji heterogenní, jinak homogenní. Ukázka schématu heterogenní informační sítě se třemi typy hran a dvěma typy vrcholů je na obrázku 11. V této síti se tedy mohou objevit hrany typu *sleduje* mezi



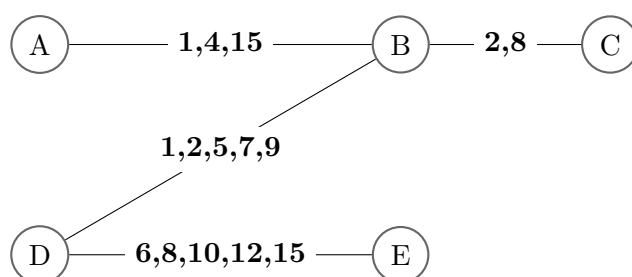
Obrázek 10: Two-mode síť



Obrázek 11: Schéma heterogenní informační sítě pro sociální síť Twitter

dvěma vrcholy typu *user*, hrany typu *zveřejňuje* mezi vrcholem typu *user* a vrcholem typu *tweet* a hrany typu *zmiňuje* mezi *tweet* a *user*. Pomocí tohoto schématu lze snadno formulovat vzory, například vzor  $user1 \rightarrow user2 \rightarrow tweet$  znázorňuje uživatele sledujícího jiného uživatele, který zveřejnil tweet [1].

### 3.2.4 Temporální síť



Obrázek 12: Temporální síť (sekvence kontaktů)

Temporální síť, někdy také označovány jako v čase se měnící síť, jsou sítě, ve kterých jsou hrany aktivní pouze v určitém časovém okamžiku nebo časovém intervalu [10]. Čas, kdy je hrana aktivní je explicitně uveden u specifikace sítě. Na obrázku 12 je znázorněna jednoduchá temporální síť, můžeme si všimnout, že jsou zde celkem 4 hrany. Každá hrana má u sebe sekvenci

čísel, tyto čísla označují čas, kdy je daná hrana aktivní. V čase 1 jsou tedy aktivní pouze hrany  $(A, B)$  a  $(B, D)$ , v čase 2 hrany  $(B, C)$  a  $(B, D)$  atp.

Topologie temporálních sítí se neustále mění, podle toho v jakém z časových intervalů na síť pohlédneme. Některé hrany zaniknou, jiné se zase objeví. Temporální síť lze rozdělit do časových oken, podle toho v kolika časových okamžicích se změní jejich topologie a následně je můžeme analyzovat staticky, jak jsme u sítí zvyklí. Kdybychom měli temporální síť přirovnat k vícevrstevným, tak si lze každé z časových oken představit jako jednu vrstvu vícevrstvé sítě. Temporální síť na obrázku 12 bychom mohli rozdělit do několika časových oken, a příslušná vícevrstvá síť by každé časové okno reprezentovala pomocí jedné vrstvy.

## 4 Vlastnosti vícevrstvých sociálních sítí

V této kapitole se blíže podíváme na míry, které dokáží kvantitativně popsat vícevrstvé sítě. Některé z uvedených vlastností mají svůj protějšek u jednovrstvých sítí, jiné jsou pak specifické pro vícevrstvé sítě.

Nejprve se zaměříme na míry aktérů, které popisují jejich vlastnosti v závislosti na jejich propojení na různých vrstvách. Některé míry jsou pouze rozšířením například **degree** centrality, **betweenness** centrality a shlukovacího koeficientu, se kterými jsme se seznámili v kapitole 2.3.4. Jiné jsou pak specifické pro vícevrstvé sítě, jako například **relevance** aktéra v dané vrstvě. Míry vrstev se pak zaměřují na vztahy mezi různými vrstvami, například jejich **similarity** (podobnost).

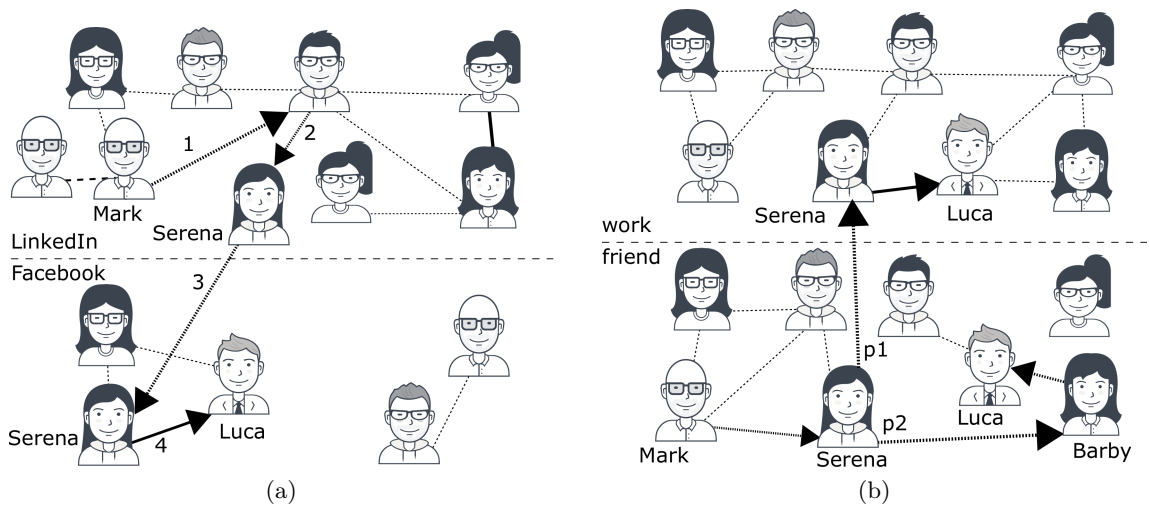
### 4.1 Čtyři hlavní přístupy

Existence interlayer hran ve vícevrstvých sítích nabízí několik přístupů, které při jejich analýze můžeme zvolit. Výsledný přístup, který pro analýzu zvolíme závisí primárně na cíli analýzy.

**První z přístupů** spočívá ve spojení všech vrstev vícevrstvé sítě do jedné, výsledkem pak je běžná jednovrstvá síť. Tento proces se nazývá **flattening** a je několik možností jak jej provést. Nejjednodušším způsobem je vytvořit novou síť, kde pro každého aktéra vícevrstvé sítě bude vytvořen korespondující vrchol a pokud alespoň na jedné z vrstev existuje hrana mezi dvěma aktéry, tak vložíme hranu i mezi korespondující vrcholy v nově vytvořené síti [1]. Pro zachování více informací o vazbách mezi aktéry, můžeme hranám v nově vytvořené síti přidat váhu, která bude rovna celkovému počtu hran mezi těmito aktéry napříč všemi vrstvami. Proces flatteningu je pak podrobněji popsán v kapitole 4.3. Výhodou tohoto přístupu je možnost využití mír, které známe z jednovrstvých sítí.

**Druhý přístup** pouze spočítá požadované míry samostatně pro každou z vrstev [1]. Vždy se vybere požadovaná vrstva, která představuje běžnou jednovrstvou síť a následně se provede její analýza. Výsledky pak lze porovnat mezi jednotlivými vrstvami, vzít například maximální nebo průměrné hodnoty, to už závisí na cíli analýzy. Tento přístup se obvykle provádí společně s flatteningem, který při spojení vrstev může některé informace znehodnotit. Předchozí dva přístupy byly založeny na aplikování postupů, které běžně používáme u jednovrstvých sítí na flattened síť, nebo na jednotlivé vrstvy. Třetí a čtvrtý z přístupů se snaží vícevrstvé sítě analyzovat bez nutnosti modifikace původní sítě.

**Třetí přístup** sice do výpočtů zahrnuje existenci několika vrstev, nicméně tyto vrstvy jsou zde považovány za stejné (hrana na jedné vrstvě má stejnou váhu jako na jiné, i když se jedná o úplně odlišné vrstvy) [1]. U metod založených na tomto přístupu je ovšem možno explicitně specifikovat odlišnost mezi interlayer a intralayer hranou, nebo dokonce mezi jednotlivými vrstvami. Toto je většinou realizováno pomocí váhy hrany, kde interlayer hrany budou mít jinou váhu, než intralayer a také hrany na jednotlivých vrstvách budou mít svou váhu, která může reprezentovat například náročnost komunikace na dané vrstvě. Pro někoho může být jednodušší napsat ka-



Obrázek 13: (a) Náhodná procházka se čtyřmi kroky, kde aktér Serena změní vrstvu LinkedIn na Facebook, jelikož je přítomna na obou z nich, (b) Dvě alternativní cesty  $p_1$  a  $p_2$  od Marka k Luca [1]

marádovi zprávu na online sociální síti, než ho osobně kontaktovat. Výsledné metody, založené na tomto přístupu, pak pro výpočet obvykle nějakým způsobem využívají náhodnou procházku, která náhodně prochází mezi jednotlivými aktéry a je také schopna procházet mezi různými vrstvami [18]. Na obrázku 13a pak můžeme vidět náhodnou procházku, která začala na vrstvě LinkedIn, konkrétně na uživateli Mark. Každý krok je pak znázorněn pořadovým číslem (1 – 4) a čárkovanou šipkou. Nejprve byl tedy náhodně zvolen Markův soused Bin ( $Mark \rightarrow Bin$ ). Následně je opět náhodně zvolen Binův soused ( $Bin \rightarrow Serena$ ). Jelikož se jedná o přístup, který zahrnuje všechny vrstvy, Serena se rozhodne změnit svou vrstvu, přesune se tedy z vrstvy LinkedIn na Facebook. Na této vrstvě náhodná procházka skončí, když Serena kontaktuje Lucu ( $Serena \rightarrow Luca$ ) [1].

**Pomocí čtvrtého přístupu** jsme schopni, stejně jako u třetího, analyzovat celou vícevrstvou síť, nicméně tento přístup považuje hrany z různých vrstev za neporovnatelné. Když se podíváme na obrázek 13b, vidíme zde dvě cesty:

$$\begin{aligned}
 P1 &= Mark_{friend} \rightarrow Serena_{friend} \rightarrow Serena_{work} \rightarrow Luca_{work} \\
 P2 &= Mark_{friend} \rightarrow Serena_{friend} \rightarrow Barby_{friend} \rightarrow Luca_{friend}.
 \end{aligned}$$

V rámci tohoto přístupu tyto cesty tedy nelze porovnat, nedává zde smysl, aby jedna cesta, která obsahuje úplně jiný typ hrany byla považována za kratší [1].

## 4.2 Míry aktérů

V této kapitole se zaměříme na vlastnosti aktérů. S některými z představených vlastností jsme se už seznámili v kapitole 2.3.4, nicméně každá z vlastností představených v této kapitole bude

Tabulka 1: Degree centrality aktérů na specifických vrstvách dle obrázku 1

actor	Facebook	LinkedIn	work	friend	flattened
Mat	1	2	3	3	9
Mark	1	2	2	3	8
Bin	0	5	3	1	9
Luca	2	0	3	2	7
Barby	0	2	2	2	6
Stine	0	2	3	0	5
Cici	2	2	2	2	8
Serena	2	1	2	3	8

upravená tak, aby ji bylo možné použít v rámci vícevrstevných sítí.

#### 4.2.1 Míry založené na stupni

Pro výpočet měr založených na **stupni** aktéra je nezbytné znát jeho stupeň, nicméně u vícevrstevných sítí se stupeň aktérů počítá pomocí definice 4.1, která stupeň počítá přes vybranou množinu vrstev sítě. Pokud ovšem pracujeme s flattened sítí, stupeň aktéra zde budeme počítat stejně jako u jednovrstevných sítí.

**Definice 4.1** *Degree centralita* aktéra  $\alpha \in A$  ve vícevrstvé síti  $G$  je rovna počtu jeho vazeb na množině vrstev  $l \subseteq L$ , kde  $G = (A, L, V, E)$  [1].

Pokud tedy chceme vypočítat stupeň aktéra ve vícevrstvé síti, projdeme všechny vrstvy na kterých nás stupeň zajímá a sečteme vazby daného aktéra na každé z těchto vrstev. V extrémním případě, kdy  $l = L$  (hledáme stupeň aktéra napříč všemi vrstvami), je stupeň aktéra stejný jako v případě flattened sítě.

V tabulce 1 pak můžeme vidět **degree centrality** všech aktérů na všech vrstvách z úvodního příkladu 1. Pokud se například rozhodneme porovnat offline a online vrstvy zjistíme, že offline vrstvy mají celkem 36 vazeb, zatímco online pouze 24. Můžeme se také podívat na nejdůležitější aktéry jednotlivých vrstev z pohledu degree centrality:

**Facebook** – Luca(2), Cici(2), Serena(2),

**LinkedIn** – Bin(5),

**work** – Mat(3), Bin(3), Luca(3), Stine(3),

**friend** – Mat(3), Mark(3), Serena(3)

Na sociální síti LinkedIn je dle degree centrality nejdůležitějším aktérem Bin, v rámci ostatních vrstev jsou ovšem degree centrality vyrovnanější a je zde několik aktérů s nejvyšší hodnotou v rámci dané vrstvy. Může nás také zajímat aktér s nejvyšší degree centralitou v celé síti, což

zjistíme sečtením všech hodnot v řádku tabulky:

**Mat** – 9, **Mark** – 8, **Bin** – 9, **Luca** – 7,

**Barby** – 6, **Stine** – 5, **Cici** – 8, **Serena** – 8

Po sečtení hodnot každého řádku zjistíme, že v rámci všech vrstev mají nejvyšší degree centralitu Mat a Bin s hodnotou 9. I takto triviální centralitu, jako je degree, můžeme ve vícevrstvých sítích použít několika způsoby a téměř vždy získat zajímavé výsledky. Při výsledné analýze bychom se tedy určitě měli podívat na výsledky jak v rámci jednotlivých vrstev, tak i celé sítě a z těchto výsledků pak zvolit, který pro nás lépe vystihuje důležitost aktérů.

Degree centralita patří mezi jednu z nejběžněji používaných a sama o sobě téměř nic nového nepřináší. Na základě této centrality je ovšem postavena **degree deviation** míra, která je specifická pro vícevrstvé sítě a poskytuje zajímavý pohled na aktivitu aktérů v jiných vrstvách. Pomocí této míry můžeme zjistit, jak moc se přítomnost daného aktéra liší v rámci vybraných vrstev [1].

**Definice 4.2** *Degree deviation* aktéra  $\alpha \in A$  ve vícevrstvé síti  $G$  je rovna směrodatné odchylce stupňů aktéra  $\alpha$  na množině vrstev  $l \subseteq L$ , kde  $G = (A, L, V, E)$  [1].

V tabulce 1 jsme si mohli všimnout, že Bin má na vrstvě LinkedIn mnohem vyšší hodnotu degree centrality, než na jakékoli jiné. Pokud bychom chtěli spočítat jeho degree deviation napříč všemi vrstvami, která je definována pomocí definice 4.2, tak bychom nejdříve zjistili jeho průměrný stupeň:

$$\frac{0 + 5 + 3 + 1}{4} = 2,25$$

a následně samotnou degree deviation:

$$\sqrt{\frac{(0 - 2,25)^2 + (5 - 2,25)^2 + (3 - 2,25)^2 + (1 - 2,25)^2}{4}} = 3,6875$$

Vysoká hodnota degree deviation pak značí, velice odlišné chování aktéra na různých vrstvách, na některých vrstvách nemusí být téměř vůbec aktivní, zatímco na jiných je jeho aktivita podstatně vyšší. Jelikož má Bin podstatně vyšší hodnoty degree centrality na vrstvách spojených s prací, než například na vrstvě Facebook, nebo vrstvě friend, můžeme předpokládat, že Bin je někdo s důležitou pracovní pozicí.



### 4.2.2 Míry založené na sousedství

Další z vlastností, pomocí kterých lze důležitost aktérů určit, je **neighborhood centralita**. K výpočtu této centrality ovšem musíme znát sousedy aktéra na vybrané množině vrstev, které získáme pomocí definice 4.3.

**Definice 4.3** *Sousedé (**neighbors**) aktéra  $\alpha \in A$  ve vícevrstvé síti  $G$  jsou různí aktéři, kteří sousedí s aktérem  $\alpha$  na množině vrstev  $l \subseteq L$ , kde  $G = (A, L, V, E)$  [1].*

Pokud se tedy opět vrátíme k úvodnímu příkladu 1 a podíváme se na sousedy Marka na vrstvách work a friend, tak si můžeme všimnout, že na vrstvě work má sousedy Mata a Cici, na vrstvě friend má včetně předchozích navíc sousedku Serenu. Jeho sousedé jsou tedy Mat, Cici a Serena. Neighborhood centralita pak pouze značí kolik sousedů daný aktér na vybrané množině vrstev má.

**Definice 4.4** *Neighborhood centralita aktéra  $\alpha \in A$  ve vícevrstvé síti  $G$  je počet různých aktérů, kteří sousedí s aktérem  $\alpha$  na množině vrstev  $l \subseteq L$ , kde  $G = (A, L, V, E)$  [1].*

Neighborhood centralita Mata na vrstvách work a friend by pak byla 3, jelikož na těchto dvou vrstvách měl tři různé sousedy. Jeho degree centralita by ovšem byla 5, jelikož na těchto dvou vrstvách má celkem pět vazeb. Degree centralita nebere v potaz, jestli se jedná o vazby se stejnými aktéry, jako na jiných vrstvách. Pokud jsou aktéři spojení vazbou se stejnými aktéry napříč různými vrstvami mají pak vysokou hodnotu **connective redundancy**. Tuto hodnotu lze získat jednoduše poměrem neighborhood a degree centrality 4.5.

**Definice 4.5** *Connective redundancy aktéra  $\alpha \in A$  ve vícevrstvé síti  $G$  je poměr mezi jeho neighborhood a degree centralitou na množině vrstev  $l \subseteq L$ , kde  $G = (A, L, V, E)$  [1].*

Connective redundancy Mata na vrstvách work a friend bychom tedy vypočítali následovně:

$$1 - \frac{\text{neighborhood}(\text{Mat}, \{\text{work}, \text{friend}\})}{\text{degree}(\text{Mat}, \{\text{work}, \text{friend}\})} = 1 - \frac{3}{5} = 0,4$$

Pokud je connective redundancy rovna 0, tak je každá vazba na každé vrstvě potřebná k tomu, aby si daný aktér udržel svou konektivitu, jelikož s žádnými aktéry nemá vazbu napříč několika vrstvami.

Míra **exclusive neighborhood** opět pracuje se sousedy daného aktéra, definována je pak pomocí 4.6.

**Definice 4.6** *Exclusive neighborhood aktéra  $\alpha \in A$  ve vícevrstvé síti  $G$  je počet jeho sousedů na množině vrstev  $l \subseteq L$  bez sousedů na  $L \setminus l$ , kde  $G = (A, L, V, E)$  [1].*

Pokud hledáme exclusive neighborhood daného aktéra na nějaké množině vrstev, tak nejprve najdeme jeho sousedy na této množině vrstev a z těchto sousedů odebereme všechny, se kterými je sousedem i na jiných vrstvách kromě dané množiny.

Pro lepší představu se opět můžeme podívat na úvodní příklad 1, konkrétně na aktéra Bina. Pokud si spočítáme jeho exclusive neighborhood na vrstvách LinkedIn a work, vyjde nám 5. Tato hodnota značí relativně vysokou exkluzivitu sousedů na vrstvách LinkedIn a work (Bin tyto aktéry nemá za sousedy na jiných vrstvách). Tento výsledek dále může naznačovat, že Bin je s těmito aktéry ve spojení výhradně kvůli práci.

### 4.2.3 Rozšíření degree centrality

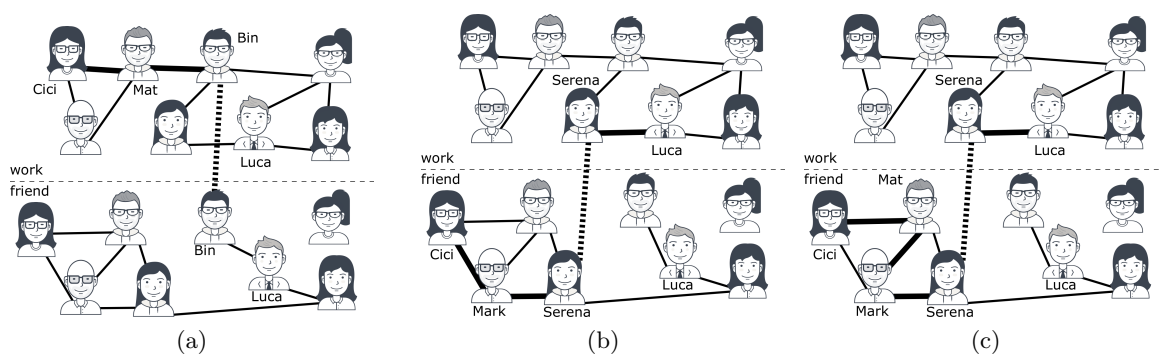
V kapitole 4.1 jsme se zmínili o možnosti použití náhodné procházky při analýze vícevrstevných sítí. První z vlastností, o kterých se zde zmíníme a které využívají náhodnou procházku se nazývá **occupation centralita**. Tuto vlastnost lze do jisté míry považovat za rozšíření degree centrality, definována je pak pomocí definice 4.7.

**Definice 4.7** *Occupation centralita* aktéra  $\alpha \in A$  ve vícevrstvé síti  $G = (A, L, V, E)$  je pravděpodobnost završení náhodné procházky sítí  $G$ , na libovolném vrcholu, který koresponduje s aktérem  $\alpha$  [1].

Opět si ukážeme příklad na vícevrstvé síti z obrázku 1. Náhodnou procházku zahájíme na aktérovi Mat na vrstvě Facebook. V každém kroku se můžeme přesunout k některému ze sousedů na aktuální vrstvě (pravděpodobnost výběru sousedů má rovnoměrné rozdělení), nebo se můžeme přesunout na úplně jinou vrstvu, pokud se Mat na některé vyskytuje. S malou pravděpodobností je také možné vybrat úplně nového aktéra. Výběr nového aktéra zabezpečuje nezastavení náhodné procházky, pokud nelze provést žádný z předchozích kroků. Mat má na Facebooku pouze jednoho přítele, ale vyskytuje se na všech čtyřech vrstvách. Při náhodné procházce je každý z těchto kroků vždy zvolen náhodně. Řekněme tedy, že zde bylo náhodně zvoleno přesunutí na vrstvu work. A posledním krokem pak bylo přesunutí k aktérovi Cici. Při náhodné procházce jsme tedy udělali celkem tři kroky a dvakrát jsme navštívili Mata a jednou Cici. Na základě těchto výsledků by occupation centralita Mata byla  $\frac{2}{3}$  a Cici pak  $\frac{1}{3}$ . Při takto malém počtu kroků je první zvolený aktér značně zvýhodněn, nicméně při větším počtu kroků tato výhoda zaniká. U ohodnocených sítí je pak možné modifikovat pravděpodobnost výběru dané hrany její váhou.

### 4.2.4 Vícevrstvé cesty

V kapitole 2.3.4 jsme se seznámili se třemi hlavními centralitami, pomocí kterých lze v sítích měřit důležitost jednotlivých vrcholů. Zatímco degree centralita určuje význam aktérů podle počtu jejich sousedů, centrality closeness a betweenness k výpočtu využívají nejkratší cesty.



Obrázek 14: Tři různé cesty mezi Cici a Lucou; (a) a (b) mají stejný počet kroků, ale jiné délky na jiných vrstvách, (c) je delší než (b) [1].

Délku cesty zde budeme chápat jako počet hran, přes které je nutno projít, abychom se dostali do cílového vrcholu.

U vícevrstvých sítí vzdálenost mezi dvěma aktéry nabývá nových rozměrů. Měřit vzdálenost mezi jednotlivými aktéry pomocí nejkratších cest, jako v případě jednovrstvých sítí, už nemusí být ideální volbou. Ve vícevrstvých sítích se aktéři často nacházejí na několika vrstvách, ve kterých se mohou libovolně pohybovat. Místo volby nejkratší cesty, může aktér zvolit cestu delší na jiné vrstvě, která mu může více vyhovovat [1]. Pro lepší představu se podívejme na vícevrstvou síť z obrázku 1. Mat je na dovolené a chtěl by kontaktovat někoho ze svých přátel. Jelikož je na dovolené, tak nejspíše nemůže využít offline vrstvu friend. Pokud by například chtěl kontaktovat Lucu, může použít Facebook. Kdyby ovšem chtěl kontaktovat Serenu, se kterou není ve spojení na žádné online vrstvě, musel by využít nějakého prostředníka. Ideálním prostředníkem by nejspíše byl opět Luca, jelikož je s Matem ve spojení na Facebooku a zároveň je přítel Sereny.

Nejkratší cesty nemusí být vždy ty, které aktéři vícevrstvých sítí zvolí, jelikož často mají na výběr spoustu jiných možností, které sice nemusí být nejkratší, ale v dané situaci mohou aktérovi více vyhovovat.

I přesto, že closeness i betweenness centrality měří vzdálenosti mezi dvěma aktéry pomocí nejkratších cest, tak z předchozího příkladu jasně vidíme, zejména u vícevrstvých sítí, že ne vždy by nás měly zajímat pouze nejkratší cesty. Existuje mnoho vlastností, které měří vzdálenosti mezi dvěma aktéry alternativním způsobem. Mezi tyto míry patří například rozšíření closeness a betweenness centralit pomocí náhodné procházky. V rámci vícevrstvých sítí je velice obtížné definovat, co je to nejkratší cesta, jelikož bychom museli být schopni nějakým způsobem porovnat vztahy na různých vrstvách a také se vypořádat se situací, kdy se aktér jednoduše rozhodne změnit svou vrstvu [1].

Jednou z možností, jak délku cesty ve vícevrstvé síti spočítat, je pomocí flatteningu, který nám umožní délky spočítat stejně jako v případě jednovrstvých sítí. Pokud bychom ovšem chtěli zachovat informaci o počtu kroků na různých vrstvách, bylo by vhodné použít vícevrstvou délku cesty 4.8. Na obrázku 14 můžeme vidět tři různé cesty. Cesty (a) a (b) mají stejný počet kroků,

Tabulka 2: Nejkratší vícevrstvé cesty mezi Stine a Markem

Cesta	LinkedIn	work	friend	Facebook
1	2			
2		3		
3	1		2	
4		2		1
5	1	2		
6		2	1	
7		1	2	
8		1	1	1

jejich vzdálenost je tedy čtyři. Nicméně, počty kroků na vrstvě work a friend se liší, cesta (a) je složena ze dvou kroků na vrstvě work, změny vrstvy a závěrečného kroku na vrstvě friend. Cesta (b) má ale dva kroky na vrstvě friend, změnu vrstvy a poslední krok na vrstvě work. V knize [1] byla pro reprezentaci délky vícevrstvé cesty navržena matice (viz definice 4.8).

**Definice 4.8** *Vícevrstvá délka cesty  $p$  na vrstvách  $L = \{l_1, \dots, l_m\}$  je matice  $M$ , kde  $M_{l_i, l_j}$  je počet hran, které je z vrcholu na vrstvě  $l_i$  k vrcholu na vrstvě  $l_j$  nutno projít [1].*

Pokud bychom cesty (a) a (b) z obrázku 14 vyjádřili pomocí definice 4.8 vypadalo by to následovně:

$$\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix} \quad (13)$$

Hodnota na prvním řádku a prvním sloupci označuje počet kroků na vrstvě work, na druhém řádku a druhém sloupci je pak počet kroků na vrstvě friend. Hodnota 1 na prvním řádku a druhém sloupci pak označuje změnu vrstvy, v případě druhé matice je to druhý řádek a první sloupec, jelikož Cici přešla z vrstvy friend na vrstvu work. Pokud bychom se rozhodli změny vrstev do délky cesty nezapočítat, mohli bychom předchozí dvě cesty vyjádřit pomocí vektorů:

$$\begin{pmatrix} 2 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \end{pmatrix} \quad (14)$$

Pokud bychom kdykoli potřebovali s délkou cesty pracovat jako s jednou číselnou hodnotou, pouze sečteme všechny složky vektoru. V tomto případě by obě cesty měly délku tři.

Jak můžeme vidět z předchozího příkladu, vícevrstvá délka cesty umožňuje uchovat více informací o vzdálenosti dvou aktérů a zároveň neztrácí možnost vyjádření délky cesty pomocí jedné číselné hodnoty, jak jsme u jednovrstvých sítí zvyklí. Na obrázku 14 jsme měli možnost vidět celkem tři vícevrstvé cesty. Cesty (a) a (b) jsou neporovnatelné, jelikož se liší v počtech kroků na různých vrstvách. Cesta (c) je ovšem nepochybně delší než cesta (b), jelikož má stejný počet kroků na vrstvě work, ale o jeden krok více na vrstvě friend [1].

V tabulce 2 můžeme vidět souhrn nejkratších vícevrstevných cest, vyjádřených pomocí zjednodušené vektorové reprezentace, mezi Stine a Markem z úvodního obrázku 1. Dle této tabulky má Stine na výběr spoustu možností, které může pro kontaktování Marka zvolit. Může zvolit nejkratší a nejjednodušší vrstvu LinkedIn, a pomocí jednoho prostředníka Marka kontaktovat, na vrstvě work ho pak může kontaktovat už pomocí tří prostředníků. Ten, koho Stine kontaktuje, ovšem může libovolně změnit svou vrstvu, může například jít na Facebook a tam požádat svého kamaráda, který se s Markem zná na vrstvě friend, aby mu zprávu předal. U ohodnocených sítí pak lze místo započítávání jednotlivých hran počítat s vahou dané hrany. Délka cesty by pak byla rovna součtu vah všech hran na dané cestě.

#### 4.2.5 Vícevrstvá closeness centralita

S closeness centralitou jsme se seznámili v kapitole 2.3.4, řekli jsme si, že jedná o jeden z možných způsobů, jakým měřit důležitost vrcholů v síti. Closeness ke stanovení důležitosti vrcholů využívá nejkratších cest, samotná closeness vrcholu je pak inverzí průměrné vzdálenosti od všech ostatních vrcholů v síti.

Stejně jako u všech předchozích vlastností, tak i u closeness bychom mohli pouze provést flattening celé sítě a následně použít běžný výpočet. Zajímavějším řešením pak je využití náhodné procházky k výpočtu closeness centrality.

**Definice 4.9** *Random walk closeness centralita* aktéra  $\alpha$  je inverzní hodnotou průměrného počtu kroků náhodné procházky začínající v libovolném jiném aktérovi potřebných k prvnímu dosažení aktéra  $\alpha$  [1].

#### 4.2.6 Vícevrstvá betweenness centralita

Stejně jako v případě vícevrstvé closeness centrality, tak i pro výpočet vícevrstvé betweenness lze využít náhodnou procházku.

**Definice 4.10** *Random walk betweenness centralita* aktéra  $\alpha$  je průměrem počtu náhodných procházek ze všech možných počátečních vrstev, mezi libovolnou dvojicí vrcholů, které prochází jakýmkoli vrcholem korespondujícím s aktérem  $\alpha$  [1].

Betweenness centralita, stejně jako closeness, k výpočtu využívá nejkratší cesty. Na téma betweenness centrality bylo provedeno mnoho experimentů, které se snažily nějakým způsobem do výpočtu zahrnout i jiné, než nejkratší cesty, nicméně žádný z nich se neprosadil a preferovaným výpočtem stále zůstává ten uvedený v definici 12. Stejně jako v případě closeness centrality, tak i u betweenness lze využít náhodné procházky k jejímu výpočtu ve vícevrstevných sítích. V knize [1] bylo provedeno několik experimentů, které se soustředily na porovnání betweenness a random walk betweenness. Jeden z experimentů zkoumal korelaci mezi těmito dvěma variantami betweenness, konkrétně u sítě AUCS [19], kde korelační koeficient vyšel 0,72, což značí velmi

Tabulka 3: Relevance aktérů z obrázku 1 v rámci jednotlivých vrstev [1]

Aktér	LinkedIn	work	friend	Facebook
Cici	0,50	0,50	0,50	0,50
Mat	0,50	0,75	0,75	0,25
Mark	0,50	0,50	0,75	0,25
Bin	0,83	0,50	0,17	
Serena	0,17	0,33	0,50	0,33
Barby	0,50	0,50	0,50	
Stine	0,67	1,00		
Luca		0,60	0,40	0,40

pozitivní korelaci mezi těmito dvěma variantami. Nicméně nemůžeme s jistotou tvrdit, že aktér, který by měl vysokou hodnotu betweenness bude mít i vysokou hodnotu random walk betweenness a naopak. Z šesti nejlepších aktérů seřazených dle betweenness, byli pouze tři také nejlepší v rámci random walk betweenness, ale žádný z nich si nezachoval pořadí. Extrémním případem pak byl aktér, který měl druhou nejvyšší hodnotu betweenness, v rámci random walk varianty byl ale až na osmnácté pozici [1].

#### 4.2.7 Míry relevance

Vícevrstvé sítě umožňují analyzovat aktéry v rámci několika vrstev. I když každý z nás působí na několika vrstvách, je zřejmé, že každá z těchto vrstev má pro nás jiný význam. Pokud by nás zajímalo, jaký význam vrstva pro daného aktéra má v rámci celé sítě, využili bychom míru **relevance**.

**Definice 4.11** *Relevance* aktéra  $\alpha \in A$  ve vícevrstvé síti  $G$  je rovna poměru mezi neighborhood centralitou aktéra  $\alpha$  na množině vrstev  $l \subseteq L$  a neighborhood centralitou aktéra  $\alpha$  na  $L$ , kde  $G = (A, L, V, E)$  [1].

Můžeme si všimnout, že pro výpočet této míry pouze využíváme neighborhood centralitu a počítáme tedy poměr mezi počtem různých sousedů daného aktéra na vybrané vrstvě, nebo vrstvách a celkovým počtem jeho různých sousedů v celé síti. V tabulce 3 se můžeme podívat na relevance všech aktérů z obrázku 1 v rámci jednotlivých vrstev.

Pokud se na tabulku 3 podíváme pozorněji, můžeme si všimnout, že nejvyšší možnou hodnotu relevance má pouze Stine na vrstvě work. Hodnota relevance 1,00 značí, že Stine je na této vrstvě ve spojení se všemi svými sousedy této vícevrstvé sítě. Na obrázku 1 vidíme, že Stine má na vrstvě work tři sousedy Bina, Luca a Barby, na všech vrstvách má pak také celkem tři různé sousedy, výpočet její relevance na vrstvě work by tedy vypadal následovně:

$$\frac{\text{neighborhood}(\text{Stine}, \{\text{work}\})}{\text{neighborhood}(\text{Stine}, \{\text{Facebook}, \text{LinkedIn}, \text{work}, \text{friend}\})} = \frac{3}{3} = 1$$

Další variantou této míry je **exclusive relevance**, která pouze využívá poměr mezi exclusive neighborhood centralitou na vybrané vrstvě, nebo vrstvách a neighborhood centralitou na všech vrstvách.

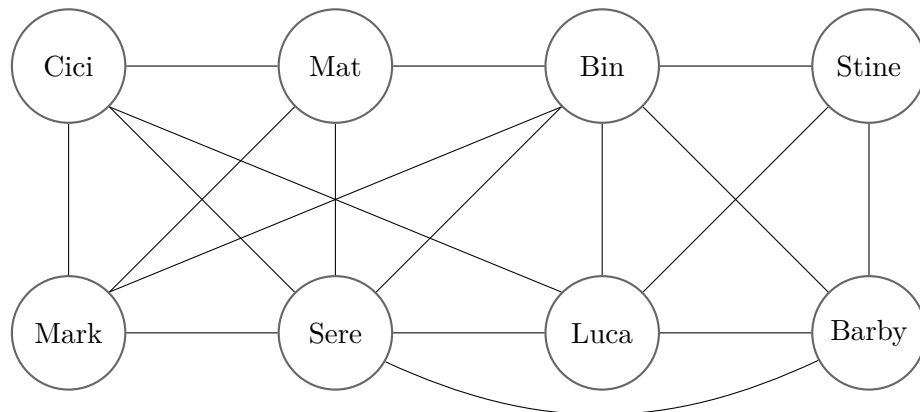
**Definice 4.12 Exclusive relevance** aktéra  $\alpha \in A$  ve vícevrstvé síti  $G$  je rovna poměru mezi exclusive neighborhood centralitou aktéra  $\alpha$  na množině vrstev  $l \subseteq L$  a neighborhood centralitou aktéra  $\alpha$  na  $L$ , kde  $G = (A, L, V, E)$  [1].

Exclusive relevanci můžeme použít například pokud bychom chtěli vědět, jaký vliv by mělo odebrání vybraných vrstev na konektivitu aktéra v celé síti. Kdybychom tedy změřili exclusive relevanci aktéra Bina na vrstvách LinkedIn a work, zjistili bychom, že je rovna 0,8. Tato vysoká hodnota exclusive relevance značí, že jeho konektivita v celé síti by byla výrazně snížena (měl by pouze jedno spojení v celé vícevrstvé síti).

### 4.3 Flattening

O procesu flatteningu jsme se již několikrát zmínili. Řekli jsme si, že se jedná o jeden z nejjednodušších přístupů, který při analýze vícevrstevných sítí můžeme zvolit. Jeho hlavní výhodou je možnost využití vlastností jednovrstevných sítí, jelikož flattened síť je jednovrstvou sítí. V této kapitole se na tento proces podíváme podrobněji.

**Definice 4.13 Flattening** vícevrstvé sítě  $G = (A, L, V, E)$  je graf  $G = (V_f, E_f)$ , kde  $V_f$  je množina aktérů, kteří se v  $G$  vyskytují alespoň na jedné vrstvě. A  $E_f$  je množina hran mezi dvojicemi aktérů sítě  $G$  [1].



Obrázek 15: Flattened síť z obrázku 1

Flattening spočívá ve vytvoření nové jednovrstvé sítě, ve které je každý aktér, který se ve vícevrstvé síti nachází alespoň na jedné vrstvě, reprezentován vrcholem. Každý z těchto aktérů, ať už se ve vícevrstvé síti vyskytuje pouze na jedné, nebo na více vrstvách, zde bude reprezentován pouze jedním vrcholem. Pokud někde ve vícevrstvé síti existuje hrana mezi dvěma aktéry, tak tato hrana bude také vložena do nově vytvořené flattened sítě mezi vrcholy reprezentující dané aktéry. Na obrázku 15 pak můžeme vidět flattened síť z úvodního příkladu 1.

Pokud bychom chtěli zachovat informaci o tom, že dva aktéři mezi sebou mají hranu napříč několika vrstvami, nabízí se zde možnost jednoduše nastavit váhu hrany v nově vytvořené jednovrstvé síti na celkový počet hran mezi těmito aktéry v původní vícevrstvé síti. Hrana mezi Cici a Markem by pak měla váhu 3, mezi Barby a Stine pak 2 atp.



## 5 Aplikace

V této kapitole se podrobněji seznámíme se samotnou aplikací, podíváme se jak na její funkční požadavky, tak na vše co je potřebné k samotnému spuštění. Následně se zaměříme na použité návrhové vzory, nejdůležitější třídy aplikace a na závěr na implementované algoritmy.

### 5.1 Specifikace

Aplikace má sloužit k analýze vícevrstvých sociálních sítí pomocí několika přístupů, se kterými jsme se seznámili v kapitole 4.1. Musí být schopna spočítat vybrané centrality, míry a metriky jak jednovrstvých sítí (na jednotlivých vrstvách a flattened sítí), tak i vícevrstvých sítí. Aplikace by také měla být schopna porovnat jednotlivé vrstvy mezi sebou. Všechno by také mělo být možné zobrazit pomocí grafů, které bude možno vyexportovat do běžně používaných formátů.

#### 5.1.1 Funkční požadavky

- Načtení vícevrstvé sítě ze souboru
- Možnost analýzy jednotlivých vrstev
- Výpočet jednovrstvých a vícevrstvých centralit, měr a metrik
- Zobrazení výsledků pomocí grafů
- Export grafů, distribucí a vypočtených výsledků do souboru

#### 5.1.2 Systémové požadavky

- MS Windows XP a vyšší
- .Net Framework 4.6.1
- Alespoň 1 GB volné operační paměti
- Vícejádrový procesor výhodou

## 5.2 Návrh

### 5.2.1 Návrhové vzory

Návrhové vzory popisují jak jednoduše a efektivně řešit situace, které se při programování opakovaně vyskytují [11].

#### Model-View View Model

Jelikož je grafické uživatelské rozhraní (GUI) implementováno pomocí Windows Presentation Foundation (WPF), byl pro oddělení logiky aplikace a uživatelského rozhraní zvolen vzor Model-View View-Model (MVVM), který je spolu s WPF velice efektivním [13].

#### Singleton

Návrhový vzor singleton použijeme pokud se chceme ujistit, že nějaká třída bude mít v celém programu pouze jedinou instanci. K této instanci pak máme přístup odkudkoli [11].

### 5.2.2 Vybraná řešení

Při návrhu aplikace se objevilo několik situací, které měly několik možných způsobů řešení. Podívejme se tedy na řešení, která nakonec byla zvolena.

#### Uživatelské rozhraní

Bylo rozhodnuto, že výsledná aplikace bude desktopovou, jako platforma byl zvolen Microsoft .Net a programovací jazyk C#. Tato platforma nabízí dvě základní možnosti, které při tvorbě uživatelského rozhraní zvolit. První možností jsou Windows Forms, což je grafická nadstavba, která pouze obaluje Windows API. Hlavní výhodou Windows Forms je, že jsou relativně starou technologií a tedy vyladěnou. Windows Forms v základu obsahují téměř jakýkoli grafický prvek, který při tvorbě uživatelských rozhraní můžeme potřebovat. Nicméně, jelikož se jedná o starší technologii, novější knihovny, zejména knihovny, které nějakým způsobem pracují s grafikou (grafy, animace, apod.), s ní nemusí být úplně kompatibilní. Druhou, novější, možností pak je WPF, což je také grafická nadstavba, která ovšem pro vykreslování využívá grafického rozhraní DirectX. Vykreslování zde probíhá vektorově, a logika aplikace je kompletně oddělena od grafického rozhraní, které je samostatně definováno v XAML (Extensible Application Markup Language) souborech. Z těchto důvodů byla pro tvorbu uživatelského rozhraní zvolena grafická nadstavba WPF.

#### Reprezentace vícevrstvé sítě

V kapitole 2.2 jsme se zmínili o třech způsobech reprezentace grafů. Nicméně každý z těchto způsobů má své výhody i nevýhody. Reprezentace vícevrstvé sítě pomocí matice sousednosti, kde

je každá vrstva reprezentována jednou maticí, má vysoké paměťové nároky, na druhou stranu některé algoritmy, např. pro výpočet délek cest, s touto reprezentací fungují velmi dobře. Pro reprezentaci vícevrstvé sítě byla tedy zvolena kombinace reprezentací pomocí matice sousednosti, seznamu sousedů a seznamu hran. Model vícevrstvé sítě obsahuje množiny aktérů, vrstev, vrcholů a hran. Každý vrchol je složen ze dvou referencí, reference na aktéra a na vrstvu. V modelu také pro každý vrchol uchováváme seznam sousedů, jelikož spousta měř vícevrstevných sítí se sousedy pracuje, a získat množinu sousedů, nebo zjistit, jestli spolu dva vrcholy sousedí v konstantním čase je velice výhodné. Každá vrstva také uchovává referenci na všechny své vrcholy i aktéry. Uchováváme pak také hrany mezi všemi dvojicemi vrstev. Takto můžeme získat všechny hrany na dané vrstvě, nebo mezi libovolnými dvěma vrstvami. Všechny vrstvy jsme pak také schopni převést do maticové reprezentace, kterou pak využíváme u některých algoritmů.

### 5.2.3 Použité knihovny

#### OxyPlot<sup>1</sup>

OxyPlot je knihovna, která umožňuje vykreslování a export grafů. Je distribuovaná pod MIT licenci [12].

#### Math.NET Numerics<sup>2</sup>

Knihovna Math.NET Numerics obsahuje spoustu matematických funkcí, které v základu v jazyce C# chybí. Tato knihovna je také distribuována pod MIT licenci.

#### Dodatečné výpočty v jazyce R<sup>3</sup>

Jazyk R s balíčkem igraph<sup>4</sup> byl použit pro výpočet betweenness centrality, která je porovnávána v kapitole 6 s random walk betweenness centralitou.

## 5.3 Implementace

Aplikace pro experimenty s vícevrstevnými sítěmi je napsaná v jazyce C#, grafické uživatelské rozhraní je implementováno pomocí WPF a jako vývojové prostředí bylo použito Visual Studio 2017.

Aplikace nabízí jednoduché uživatelské rozhraní, hlavní okno si můžeme prohlédnout na obrázku 16. Aplikace umožňuje načítat vícevrstvé sítě ze souboru, kde po načtení sítě jsou uživateli zobrazeny základní informace o aktérech, vrstvách, vrcholech a distribucích vybraných centralit. Uživatel také může porovnávat distribuce všech vrstev, včetně flattened sítě, jak můžeme

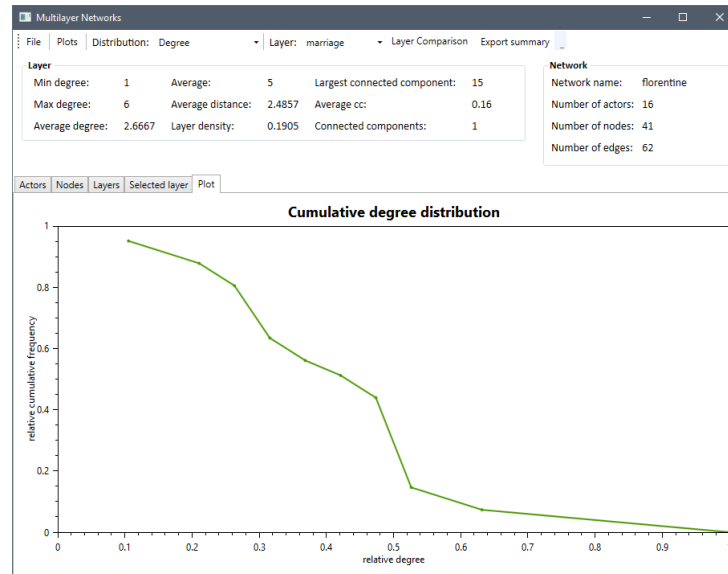
---

<sup>1</sup><http://www.oxyplot.org/>

<sup>2</sup><https://numerics.mathdotnet.com/>

<sup>3</sup><https://cran.r-project.org/>

<sup>4</sup><http://igraph.org/>



Obrázek 16: Hlavní okno aplikace

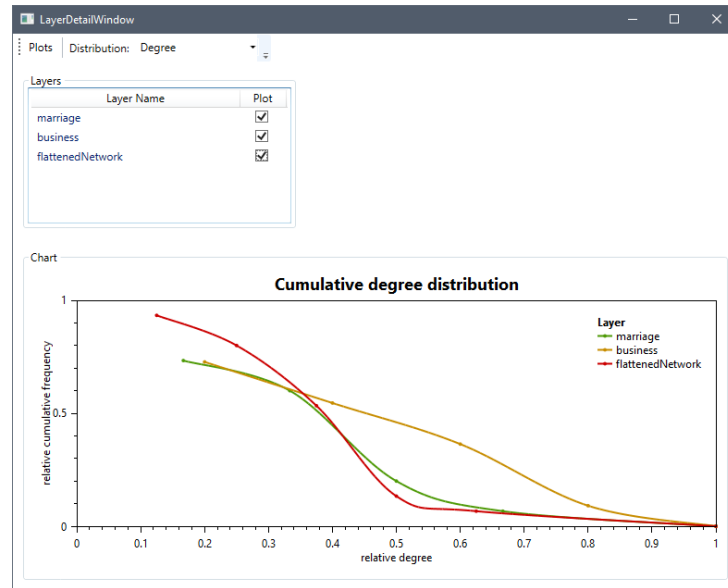
vidět na obrázku 17. Jednotlivé vrstvy lze také exportovat jako seznam hran do CSV souboru, exportovat také lze všechny grafy i souhrnnou statistiku všech vrstev.

### 5.3.1 Formát MPX

Vícevrstvé sítě jsou reprezentovány pomocí tohoto formátu. Všechny vícevrstvé sítě, se kterými jsme se v této práci setkali, byly k dispozici v tomto formátu. Každá síť v tomto formátu musí mít alespoň seznam hran, kde intralayer hrany mají formát `Actor_1, Actor_2, Layer`. Interlayer hrany pak mají formát `Actor_1, Layer_1, Actor_2, Layer_2`. Formát také umožňuje definovat několik dalších nepovinných informací, jako například atributy aktérů, vrcholů, hran nebo typ sítě.

### 5.3.2 Typické použití

Prvním krokem po spuštění aplikace je vždy načtení MPX souboru s vícevrstvou sítí (*File* → *Open*). Jakmile máme vybranou síť načtenou, zobrazí se uživateli dotazovací okno, které se ho zeptá, jestli si přeje spočítat random walk centrality, jelikož výpočet těchto centralit může trvat i na menších sítích několik minut a pokud uživatele nezajímají, není důvod je počítat. U větších sítí je pak uživatel dotázán, jestli si přeje spočítat nejkratší cesty sítě a míry s nimi spojené, následně se uživateli zobrazí informace o načtené síti. V horní části hlavního okna aplikace se nacházejí téměř všechny ovládací prvky, které slouží např. pro exportování vybrané vrstvy do CSV souboru v podobě seznamu hran nebo exportu distribucí do formátu PDF nebo PNG. Pod ovládacími prvky jsou pak informace o vybrané vrstvě, napravo od nich je pak základní popis sítě (počet aktéru, vrcholů, apod.). Ve spodní části hlavního okna jsou pak záložky s aktéry, vrcholy, vrstvami, zvolenou vrstvou a aktuálně zvolenou distribucí (viz obrázek 16).



Obrázek 17: Okno pro porovnání vrstev

Pro porovnání distribucí vrstev klikneme na tlačítko *Layer Comparison* v horní části aplikace. Otevře se nové okno, velice podobné hlavnímu (ovládací prvky v horní části, distribuce ve spodní části), pod ovládacími prvky ovšem uvidíme seznam všech vrstev včetně flattened sítě (viz obrázek 17). Vrstvy, které chceme porovnat, jednoduše vybereme ze seznamu a v horní části vybereme distribuci, kterou si přejeme napříč vybranými vrstvami porovnat. Zvolené distribuce pak můžeme, stejně jako v případě hlavního okna, vyexportovat do PDF nebo PNG formátu.

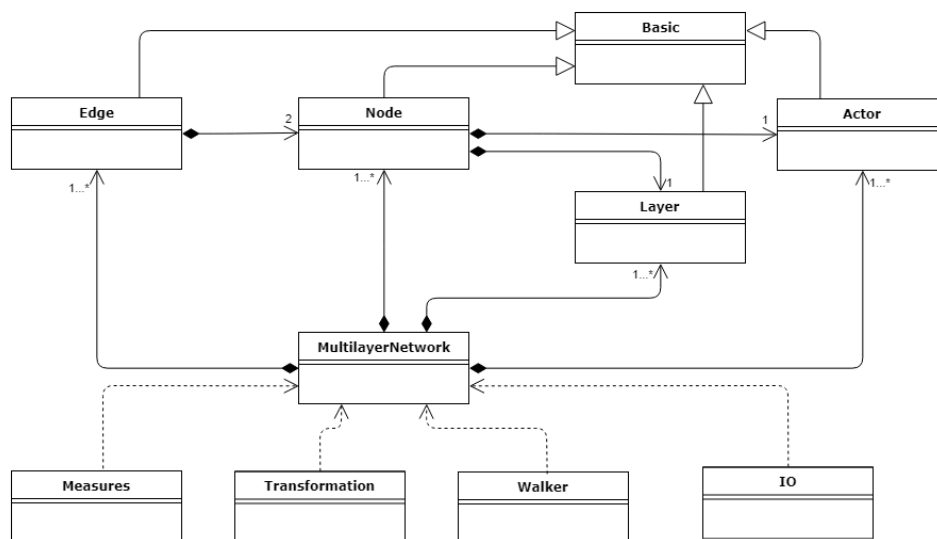
### 5.3.3 Vybrané třídy

#### BasicComponent

Třída `BasicComponent` slouží jako základní předpis pro všechny třídy, ze kterých je model vícevrstvé sociální sítě složen. Z této třídy tedy dědí třídy: `Actor`, `Node`, `Edge` a `Layer` (viz obrázek 18). Třída obsahuje pouze dva atributy: ID typu `integer` a Name typu `string`, třída pak přepisuje základní operátory jako `==`, `!=`, `<`, `>`, kde porovnání objektů pomocí těchto operátorů bude prováděno pomocí atributu ID.

#### Node

Třída `Node` reprezentuje vrchol v modelu vícevrstvé sociální sítě, kromě všech atributů třídy `BasicComponent`, obsahuje také navíc další dva atributy: objekt třídy `Actor` a `Layer` (viz obrázek 18).



Obrázek 18: Diagram tříd hlavní části aplikace

## Edge

Třída `Edge` slouží k reprezentaci hran v modelu vícevrstvé sociální sítě. Kromě atributů, která zdělila od třídy `BasicComponent`, obsahuje navíc dva atributy, které jsou objektem třídy `Node` (viz obrázek 18).

## MultilayerNetwork

Jak již název napovídá, třída `MultilayerNetwork` slouží k reprezentaci modelu vícevrstvé sociální sítě. Třída obsahuje jak atributy, které jsou už z definice vícevrstvé sítě nezbytné, tak spoustu pomocných atributů, které usnadňují práci se samotným modelem. Mezi tyto atributy patří:

- **Actors, Nodes, Edges a Layers** - hashsety objektů specifické třídy (Actors je tedy hashset objektů třídy `Actor`, atd.).
- **EmptyNodes a EmptyEdges** - pomocné hashsety, které jsou vráceny, pokud některá metoda potřebuje vrátit prázdný hashset (nemusíme vždy vytvářet nový).
- **LayerByName a ActorByName** - slovníky, ve kterých jsou klíče jména vrstev/aktérů a hodnoty jsou samotné vrstvy/aktéři (podobných slovníků je v této třídě daleko více, nicméně nebudeme je zde všechny uvádět).

Tato třída implementuje také mnoho metod, které slouží především k sestavení modelu vícevrstvé sítě. Mezi tyto metody patří například:

- **AddActor** - Z názvu je již jasné, že metoda slouží k přidání nového aktéra do modelu vícevrstvé sítě. Metoda přijímá jeden parametr `actorName` typu `string`. Prvním krokem je zjištění, jestli aktér s daným jménem již existuje, pokud ano, vrátíme tohoto aktéra, pokud ne, vytvoříme aktéra s požadovaným jménem. Nově vytvořeného aktéra následně vložíme do hashsetu s aktéry a také do slovníku `ActorByName`, kde klíčem bude jméno aktéra, a hodnotou samotný aktér. Podobné metody jsou implementovány pro všechny komponenty vícevrstvé sítě (`AddNode`, `AddEdge` a `AddLayer`). Každá z těchto metod se stará o cachování dle atributu `ID` nebo `Name` (vložení do slovníku).
- **GetNode** - Metoda přijímá dva parametry `actor` typu `Actor` a `layer` typu `Layer`. Jelikož jsou všechny komponenty cachovány, jednoduše se podíváme do slovníku `NodeByActorAndLayer` (kde klíčem je `ID` aktéra a hodnotou je další slovník, ve kterém je klíčem `ID` vrstvy a hodnotou je vrchol). Nejprve zjistíme, jestli slovník obsahuje `ID` daného aktéra. Pokud slovník dané `ID` obsahuje, podíváme se, jestli jeho hodnota, která je také slovníkem, obsahuje `ID` dané vrstvy. Pokud i tento slovník obsahuje hledané `ID`, tak vrátíme hodnotu, kterou je hledaný vrchol. Tyto metody jsou podobně implementovány i pro ostatní komponenty vícevrstvé sítě, jelikož implementace pomocí slovníků poskytuje konstantní složitost pro nalezení, nebo získání požadovaného prvku.
- **ToAdjacencyMatrix** - Tato metoda slouží k převodu vrstev vícevrstvé sítě do formátu matic sousednosti. Metoda přijímá pouze jeden parametr, kterým je objekt typu `Layer`, který má výchozí hodnotu `null` a pokud nabývá této hodnoty, budou vráceny matice sousednosti pro všechny vrstvy, jinak pouze pro vrstvu předanou v parametru. Jelikož je v maticích sousednosti nutné mít aktérův `ID` atribut indexovaný pomocí indexu v poli, musíme pro každou vrstvu vytvořit stejně velkou matici, která bude mít počet řádků i sloupců roven počtu aktérů vícevrstvé sítě. Tohle je také jeden z důvodů, proč jsme se reprezentaci vícevrstvé sítě pomocí matic sousednosti vyvarovali. Následně projdeme všechny vrstvy, a získáme jejich hrany. Do příslušné matice sousednosti pak vložíme jedničku do sloupce, který bude roven atributu `ID` aktéra prvního vrcholu získané hrany, a na řádek, který bude roven atributu `ID` aktéra druhého vrcholu. Pokud hrany na této vrstvě byly neorientované prohodíme sloupec s řádkem, a vložíme jedničku i zde.

## Walker

Třída `Walker` slouží k náhodné procházce vícevrstvou sítí. Konstruktor přijímá celkem tři parametry objekt typu `MultilayerNetwork`, proměnnou typu `double` a pole polí typu `double`. Prvním parametrem je model vícevrstvé sítě, po které bude náhodná procházka provedena. Druhým je pravděpodobnost provedení skoku na libovolný jiný vrchol. A třetím je matice, která reprezentuje pravděpodobnost přechodu mezi každou dvojicí vrstev, tento parametr je nepovinný a jeho výchozí hodnotou je rovnoměrné rozdělení pravděpodobnosti mezi všemi vrstvami.

## Measures

Třída **Measures** implementuje celkem 10 algoritmů, které slouží k výpočtu centralit/měr vícevrstvé sítě, které byly popsány v kapitole 6.4. Mezi metody, které tyto algoritmy implementují, patří například:

- **Degree** - Slouží k výpočtu degree centrality aktéra na vybrané množině vrstev. Přijímá celkem čtyři parametry, objekt typu **MultilayerNetwork**, **Actor**, **EdgeMode** a hashset objektů typu **Layer**. Nejprve je vybrán vrchol daného aktéra a následně jsou procházeni všichni sousedé tohoto vrcholu. Pokud se některý ze sousedů nachází na některé vrstvě z vybrané množiny, navýší se hodnota aktérový degree centrality o jedničku. Takto jsou procházeny všechny vrcholy daného aktéra.
- **Occupation** - Tato metoda slouží k výpočtu occupation centrality aktérů vícevrstvé sítě. Přijímá celkem pět parametrů, kde první tři slouží pro inicializaci náhodné procházky. Tyto parametry můžeme vidět výše u třídy **Walker**. Předposledním parametrem je proměnná typu **integer**, která určuje počet kroků. Posledním parametrem je objekt typu **Node**, který značí počáteční vrchol, ze kterého bude náhodná procházka začínat. Pomocí náhodné procházky procházíme vícevrstvou síť, a při každém kroku, kromě skoku a žádné akce (nebyl proveden žádný krok, nelze se v náhodné procházce pohnout dále, jelikož vrchol nemá žádné sousedy), zvýšíme hodnotu occupation centrality daného aktéra o jedničku. Náhodná procházka a také metoda končí, jakmile je vyčerpán počet kroků.
- **RandomWalkCloseness** - Metoda slouží k výpočtu random walk closeness centrality. Přijímá celkem tři parametry, které jsou stejné jako první tři u **Occupation** metody. Nejprve je pro každou dvojici vrcholů vypočítána vzdálenost pomocí náhodné procházky, která je rovna první náhodné procházce, která mezi těmito vrcholy byla provedena (nebere se tedy průměr z několika procházek, ani minimum). Jakmile jsou mezi všemi dvojicemi vrcholů spočítány vzdálenosti, je už closeness centralita pouze spočítána pomocí vzorce 11, kde délku nejkratší cesty nahradíme délkou náhodné procházky.

## Transformation

Třída **Transformation** implementuje metody pro transformaci vícevrstvé sítě, konkrétně flattening, který byl popsán v kapitole 4.3.



## 6 Experimenty

### 6.1 Popis datových sad

#### Florentine Families

Tento dataset popisuje vztahy 16 rodin, které žily ve Florencii okolo roku 1430. Dataset obsahuje celkem dvě vrstvy. Vrstva business obsahuje vazby, které mohou reprezentovat půjčky, úvěry a partnerství. Vrstva marriage pak obsahuje vazby, které reprezentují manželství [1]. Základní popis tohoto datasetu lze vidět v tabulce 4.

Tabulka 4: Dataset Florentine Families

Počet vrstev	Počet aktérů	Počet vrcholů	Počet hran
2	16	26	35

#### Sampson's Monastery

Průzkum, kterého se zúčastnilo celkem osm mnichů. Každý z mnichů byl požádán, aby ze čtyř dvojic vztahů, které byly tvořeny pozitivním a negativním vztahem, vybral tři. Dvojice vztahů, ze kterých bylo vybíráno, byly následující: respekt/disrespekt, líbí se/nelíbí se, pozitivní vliv/-negativní vliv a chvála/vina [1]. Tabulka 5 obsahuje základní popis tohoto datasetu.

Tabulka 5: Dataset Sampson's Monastery

Počet vrstev	Počet aktérů	Počet vrcholů	Počet hran
10	18	175	510

#### Bank Wiring

Dataset popisuje celkem 14 zaměstnanců americké společnosti Western Electric (viz tabulka 6). Zaměstnanci měli různé pracovní pozice (dva inspektori, tři svářeči a devět elektrikářů). Vrstvy popisují účast zaměstnanců při zábavných aktivitách, argumentech o otevřených oknech, přátelství, negativní chování, pomoc ostatním při práci, a kolikrát si pracovníci vyměnili úkoly [1].

Tabulka 6: Dataset Bank Wiring

Počet vrstev	Počet aktérů	Počet vrcholů	Počet hran
6	14	62	110

## Tailor Shop

Tento dataset popisuje vztahy mezi 39 zaměstnanci krejčovství (viz 7). Dataset je složen ze dvou verzí, kde každá z verzí byla vytvořena v jiném časovém okamžiku. Každá verze obsahuje 2 vrstvy, které popisují pracovní a přátelské vztahy zaměstnanců [1].

Tabulka 7: Dataset Tailor Shop

Počet vrstev	Počet aktérů	Počet vrcholů	Počet hran
4	39	150	552

## AUCS

Tato anonymizovaná data pochází z univerzity, a obsahují celkem 5 offline vrstev a online vrstev (viz 8). Vrstvy představují vztahy na facebooku, společné obědy, spoluautorství, společné trávení volného času a pracovní vztahy. Do průzkumu se zapojilo 61 zaměstnanců z celkového počtu 142. Jedná se o profesory, doktorandy, postdoktorské výzkumníky a administrativní pracovníky [1].

Tabulka 8: Dataset AUCS

Počet vrstev	Počet aktérů	Počet vrcholů	Počet hran
5	61	224	620

## 6.2 Porovnání jednotlivých vrstev

V kapitole 4.1 byly zmíněny čtyři možné přístupy, které při analýze vícevrstevných sítí můžeme zvolit. První z přístupů byl zaměřen na analýzu sítě pomocí flatteningu. Druhý se soustředil na analýzu jednotlivých vrstev. Cílem tohoto experimentu je porovnání těchto dvou přístupů a následné zhodnocení jejich výsledků. Pro každou síť budou její vlastnosti spočítány jak přes jednotlivé vrstvy, tak na flattened síti. Základní vlastnosti jako nejmenší stupeň, hustota nebo průměrný shlukovací koeficient budou vždy přehledně uvedeny v tabulce pro každou vícevrstvou síť. Distribuce pak budou vizuálně porovnány pomocí série spojnicových grafů s relativní kumulativní četností.

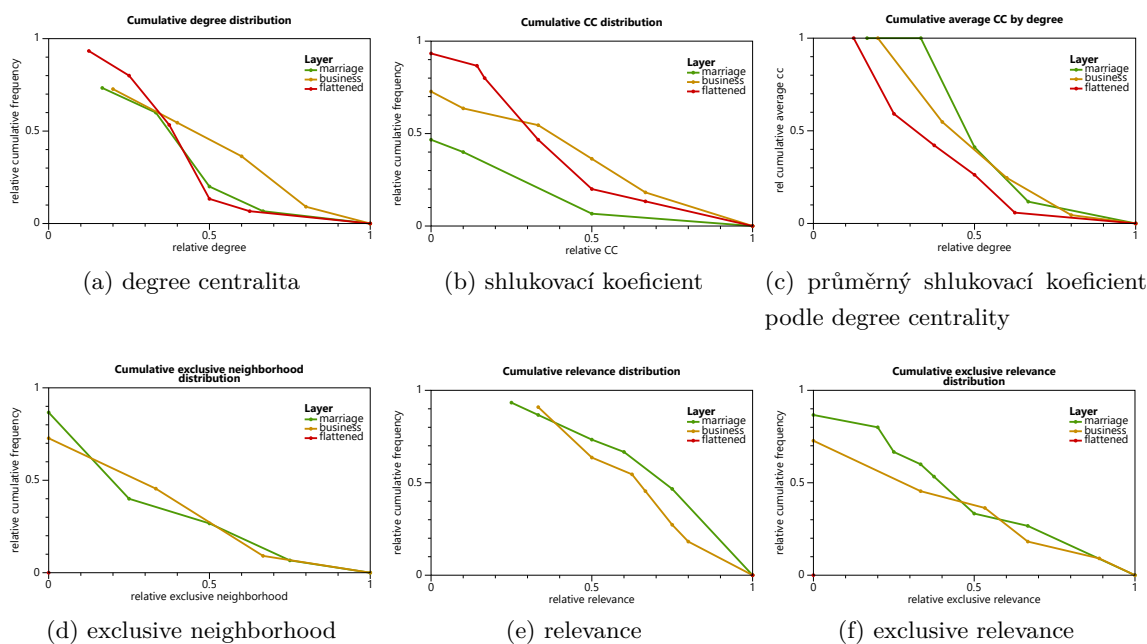
### 6.2.1 Florentine Families

V tabulce 9 jsou uvedeny základní vlastnosti jednotlivých vrstev vícevrstvé sítě Florentine Families. V posledním sloupci jsou pak vlastnosti flattened sítě. Při pohledu na vlastnosti vrstev marriage a business si můžeme všimnout, že vlastnosti, které souvisí se stupni vrcholů, si jsou velmi podobné.

Tabulka 9: Základní vlastnosti vrstev vícevrstvé sítě Florentine Families

Vrstva	marriage	business	flattened
Actors	15	11	15
Edges	20	15	27
Min. Degree	1	1	1
Max. Degree	6	5	8
Avg. Degree	2,67	2,73	3,6
Avg.Distance	2,49	2,38	2,09
Density	0,19	0,27	0,26
Avg. CC	0,16	0,43	0,44

Minimální stupně vrcholů jsou zde s hodnotou 1 stejné pro obě vrstvy, maximální stupně se pak liší minimálně. Průměrné stupně jsou taky velice podobné. Na obrázku 19e pak můžeme vidět distribuce degree centralit pro jednotlivé vrstvy, včetně flattened sítě. Z grafu můžeme vidět, že i distribuce degree centralit jednotlivých vrstev si jsou velmi podobné.



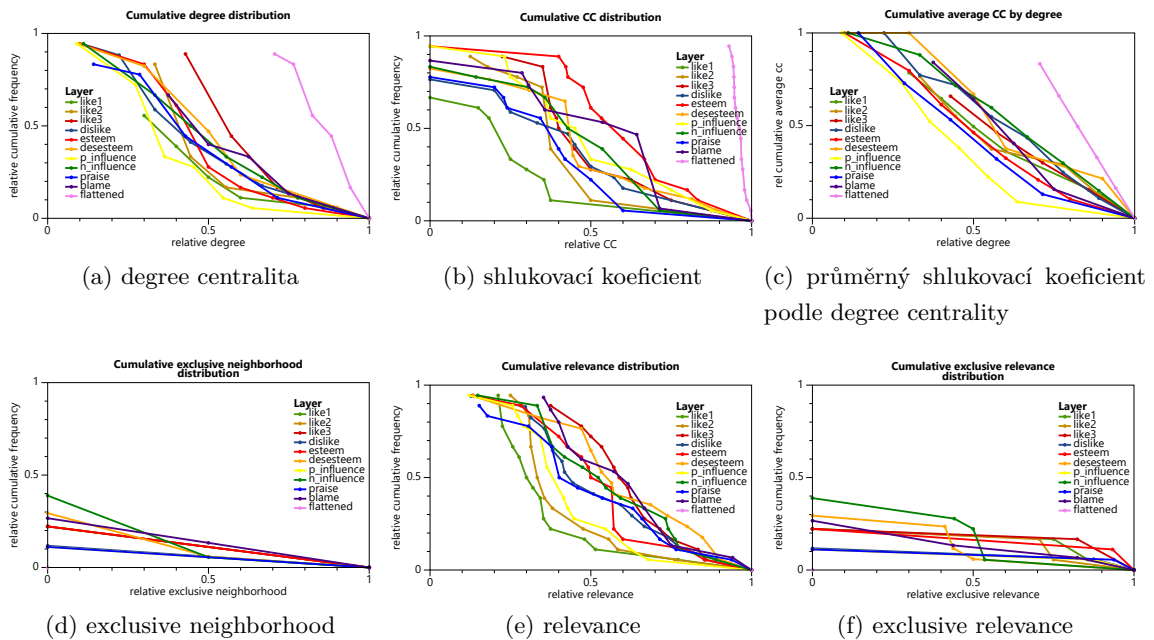
Obrázek 19: Kumulativní relativní doplňkové distribuce vrstev vícevrstvé sítě Florentine Families

Když se podíváme na flattened síť, můžeme si všimnout, že maximální stupeň se zvýšil na 6. Průměrný stupeň byl u obou vrstev velice podobný, na vrstvě marriage byl 2,67 a na vrstvě business 2,73, na flattened síti je ovšem 3,6. Tyto výsledky naznačují, že hrany mezi vrcholy na vrstvách marriage a business nebyly pouze mezi stejnými aktéry. A věci, které se týkají byznysu se neřeší pouze v rámci jedné rodiny. Průměrná vzdálenost na flattened síti je pak také nižší,

než na obou vrstvách. Toto také vyplývá z distribuce exclusive neighborhood centrality, kterou můžeme vidět na obrázku 19d, která u vrstev marriage a business obsahuje několik hodnot s vyšší hodnotou, než 0. Flattened síť pak má vždy hodnotu 0, jelikož jsou zde všichni sousedé, které aktér může ve vícevrstvé síti na různých vrstvách mít.

### 6.2.2 Sampson's Monastery

Tabulka 10 obsahuje základní vlastnosti všech vrstev vícevrstvé sítě Sampson's Monastery. Každý z aktérů na této síti musel zvolit právě tři páry pozitivních a negativních vztahů. Z tohoto důvodu mohou být vlastnosti, které k výpočtu využívají stupně aktérů, značně ovlivněny. To také můžeme vidět na obrázku 20a, kde si jsou jednotlivé distribuce degree centrality velice podobné. Pouze u flattened sítě je vidět značný rozdíl. Podle počtu hran jednotlivých vrstev můžeme usoudit, že aktéři této sítě mezi sebou mají spíše pozitivní vztahy, než negativní. Vrstvy, které představují negativní vztahy mají celkem 196 hran, pozitivní vztahy pak mají celkem 203.



Obrázek 20: Kumulativní relativní doplňkové distribuce všech vrstev vícevrstvé sítě Sampson's Monastery

Pokud se podíváme na obrázek 20e můžeme si všimnout, že vrstvy spojené s negativními vztahy, mají výrazně vyšší hodnoty relevance. To znamená, že většina aktérů, kteří nemají rádi, nebo viní tyto aktéry, s těmito aktéry na jiných vrstvách nejsou v kontaktu. Toto také potvrzuje distribuce exclusive relevance na obrázku 20f. Zde můžeme vidět, že odebráním těchto vrstev, by tito aktéři měli výrazně menší počet vazeb. Můžeme si všimnout, že vrstva praise (chvála) má nejmenší počet hran ze všech vrstev. Dočkat se chvály od mnichů může být tedy velice obtížné, alespoň v tomto chrámu. Na druhou stranu vrstva disesteem (nesouhlas) má

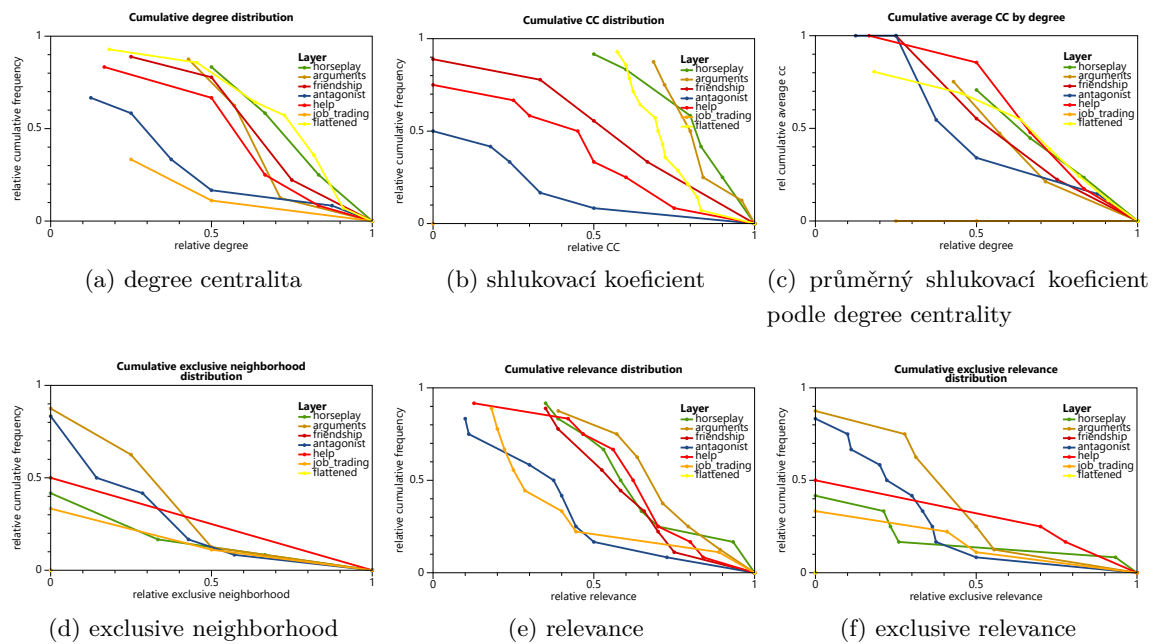
nejvyšší počet hran. U flattened sítě jsou všechny tyto informace ztraceny, jelikož zde nevíme co jednotlivé hrany reprezentují.

Tabulka 10: Vrstvy datasetu Sampson's Monastery

Vrstva	Actors	Edges	Min. Degree	Max. Degree	Avg. Degree	Avg. Dist	Density	Avg. CC
like1	18	55	3	10	4,56	2,62	0,18	0,18
like2	18	57	3	9	4,67	3,09	0,19	0,28
like3	18	56	3	7	4,56	2,91	0,18	0,33
dislike	17	47	1	9	4,47	1,44	0,17	0,14
esteem	18	54	1	10	5	1,12	0,18	0,3
disesteem	17	58	1	10	5,76	1,82	0,21	0,2
p. inf.	18	53	1	11	4,56	1,28	0,17	0,33
n. inf.	18	50	1	9	4,78	1,84	0,16	0,11
praise	18	39	1	7	3,56	1,14	0,13	0,3
blame	15	41	3	8	4,67	1,37	0,2	0,12
flattened	18	211	12	17	14,89	1,3	0,67	0,69

### 6.2.3 Bank Wiring

V tabulce 11 můžeme vidět základní vlastnosti všech vrstev vícevrstvé sítě Bank Wiring. Vrstva horseplay, která reprezentuje zábavné aktivity, je bezesporu nejpopulárnější vrstvou celé sítě. Aktéři této sítě jsou celkově aktivnější na vrstvách, které reprezentují nějaké pozitivní vztahy. To také potvrzuje distribuce degree centrality (viz obrázek 21a). Můžeme si všimnout, že vrstva arguments, která reprezentuje argumenty mezi zaměstnanci, má nejmenší počet aktérů v celé síti. Na obrázku 21e je distribuce relevance jednotlivých vrstev. Můžeme si všimnout, že právě na vrstvě arguments mají zaměstnanci nejvyšší hodnoty relevance. To znamená, že zaměstnanci, kteří s nimi mají argumenty, s nimi na jiných vrstvách v kontaktu nejsou. Podobnou situaci jsme také viděli u sítě Sampson's Monastery. Zajímavou vrstvou je také job trade, která reprezentuje výměnu úkolů mezi zaměstnanci. Nejvyšší stupeň s hodnotou 4 značí, že s některým ze zaměstnanců si 4 další vyměňují úkoly. Průměrný shlukovací koeficient je 0. Z této hodnoty můžeme usoudit, že zaměstnanci, kteří vyměňují úkoly se stejnými zaměstnanci, mezi sebou úkoly nevyměňují. U flattened sítě pak opět ztrácíme informaci o tom, co jednotlivé vrstvy reprezentují. Z výrazně vyšší hodnoty nejvyššího stupně a průměrného stupně, můžeme usoudit, že vazby na jednotlivých vrstvách nejsou vždy mezi stejnými zaměstnanci. Pokud mezi sebou někteří zaměstnanci mají argumenty, nejspíše nejsou přátelé. U této sítě si můžeme opět všimnout, že pokud jsou charakteristiky jednotlivých vrstev odlišné, flattening celé sítě není vhodným způsobem pro analýzu.



Obrázek 21: Kumulativní relativní doplňkové distribuce všech vrstev vícevrstvé sítě Bank Wiring

Tabulka 11: Základní vlastnosti vrstev vícevrstvé sítě Bank Wiring

Vrstva	horseplay	arguments	friends	antagonist	help	job trade	flattened
Actors	12	8	9	12	12	9	14
Edges	28	19	13	19	24	7	109
Min. Degree	3	3	1	1	1	1	2
Max. Degree	6	7	4	8	6	4	11
Avg. Degree	4,67	4,75	2,89	3,17	3,67	1,56	8,14
Avg.Distance	2,09	1,32	2,22	1,86	1,97	0,19	1,42
Density	0,42	0,68	0,36	0,29	0,18	0,1	0,6
Avg. CC	0,83	0,69	0,63	0,22	0,14	0	0,72

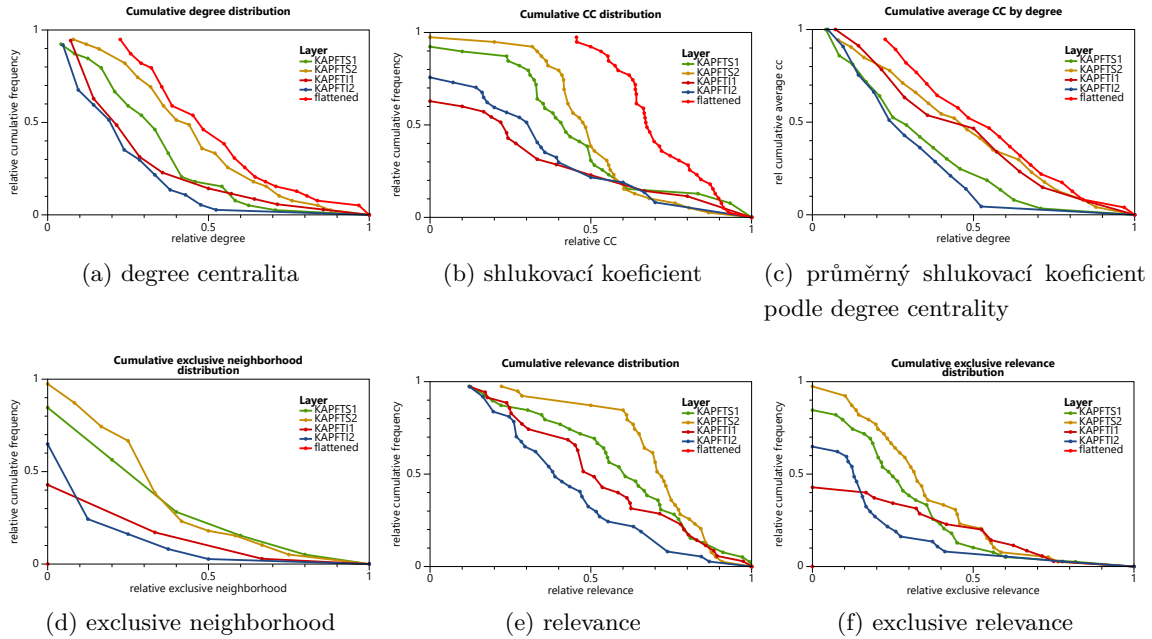
## 6.2.4 Tailor Shop

Tabulka 12 obsahuje základní vlastnosti všech vrstev vícevrstvé sítě Tailor Shop. Vrstvy KAPFTS1/2 obsahují sociální vztahy mezi zaměstnanci a vrstvy KAPFTI1/2 obsahují pracovní vztahy. Tyto dvě verze sociálních a pracovních vztahů byly pořízeny v rozmezí několika měsíců. Když se podíváme na vrstvu KAPFTS2 můžeme si všimnout, že počet hran se podstatně zvýšil. Minimální stupeň oproti KAPFTS1 také vzrostl. Každý zaměstnanec zde má už nejméně dva přátele nebo kolegy, se kterými dochází k nějakým sociálním interakcím. Distribuce degree centrality vrstvy KAPFTS2 také ukazuje, že počet zaměstnanců s vyšší hodnotou degree cen-

trality se zvýšil (viz obrázek 22a). Distribuce shlukovacího koeficientu (viz obrázek 22b) jasně ukazuje, že za těch několik měsíců se přátelé přátel začali daleko více seznamovat. Na obrázku 22e můžeme jasně vidět, že většina aktérů má vyšší hodnotu relevance právě na vrstvě KAPFTS2. Nejnížší hodnotu relevance má většina zaměstnanců na vrstvě KAPFTI2. U flattened síť můžeme vidět, že distribuce shlukovacího koeficientu je výrazně vyšší (viz obrázek 22b), než u jakékoli jiné vrstvy. Tohle může být zapříčiněno hlavně tím, že flattened síť obsahuje veškeré vztahy mezi zaměstnanci, které se zde během těch několika měsíců mohly vyskytovat. Některé vztahy, které byly na prvních verzích jednotlivých vrstev, mohly v pozdějších verzích zaniknout. Flattened síť tyto vztahy ovšem stále zachovává.

Tabulka 12: Základní vlastnosti vrstev vícevrstvé sítě Tailor Shop

Vrstva	KAPFTS1	KAPFTS2	KAPFTI1	KAPFTI2	flattened
Actors	39	39	35	37	39
Edges	158	223	76	95	312
Min. Degree	1	2	1	1	7
Max. Degree	24	25	14	21	31
Avg. Degree	8,1	11,44	4,34	5,14	16
Avg.Distance	2,04	1,77	2,51	1,86	1,58
Density	0,21	0,3	0,13	0,14	0,42
Avg. CC	0,46	0,5	0,31	0,34	0,61



Obrázek 22: Kumulativní relativní doplňkové distribuce všech vrstev vícevrstvé sítě Tailor Shop

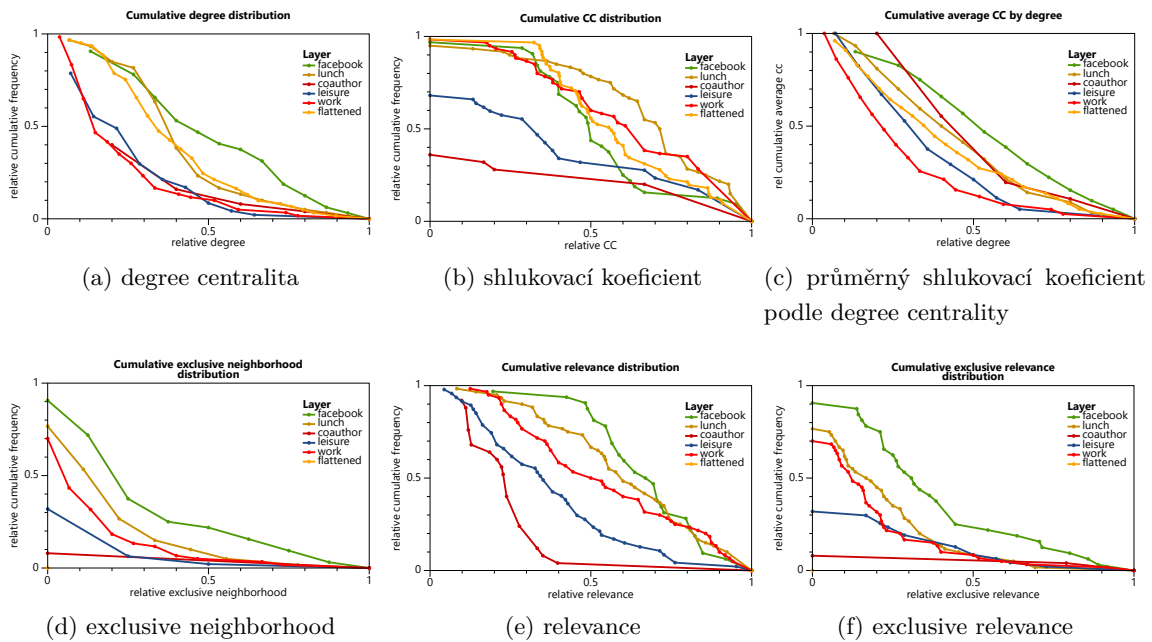
### 6.2.5 AUCS

V tabulce 13 jsou základní vlastnosti všech vrstev vícevrstvé sítě AUCS.

Tabulka 13: Základní vlastnosti vrstev vícevrstvé sítě AUCS

Vrstva	facebook	lunch	coauthor	leisure	work	flattened
Actors	32	60	25	47	60	61
Edges	124	193	21	88	194	353
Min. Degree	2	1	1	1	1	2
Max. Degree	15	15	5	14	27	29
Avg. Degree	7,75	6,43	1,68	3,74	6,47	11,57
Avg. Distance	1,96	3,19	0,19	2,74	2,39	2,06
Density	0,25	0,11	0,07	0,08	0,11	0,19
Avg. CC	0,54	0,67	0,27	0,39	0,64	0,59

Zajímavé může být, že z 60 aktérů pouze 32 využívá facebook. Po vrstvě coauthor, která reprezentuje spoluautorství, je to druhá vrstva s nejmenším počtem aktérů. Hustota vrstvy facebook je ovšem mnohem vyšší, je dokonce nejvyšší ze všech vrstev, včetně flattened sítě.



Obrázek 23: Kumulativní relativní doplňkové distribuce všech vrstev vícevrstvé sítě AUCS

Když se podíváme na distribuci shlukovacích koeficientů (viz obrázek 23b), můžeme si všimnout vysokých hodnot na vrstvách lunch a work. Což dává naprostý smysl hlavně u vrstvy lunch. Pokud aktér obědvá s jinými aktéry, tak tito aktéři spolu také automaticky musí obědvat.



Z distribuce relevance a exclusive neighborhood centrality (viz obrázek 23e) lze vidět, že jejich hodnoty jsou na vrstvě facebook výrazně vyšší, než na ostatních vrstvách. Aktéři, kteří jsou s někým přátelé na facebooku s nimi tedy většinou ve spojení na jiných vrstvách už nebývají. Toto může být zapříčiněno samotným facebookem, kde jsou lidé často stále přátelé, i když se už nevidají, například bývalí kolegové. Vysoké hodnoty relevance má taky vrstva lunch. Vrstvy této sítě si jsou vcelku podobné, a hrany na různých vrstvách si neodporují, jako například u sítě Sampson's Monastery, kde některé vrstvy reprezentovaly pozitivní vztahy a některé naopak negativní. Flattened síť by tedy měla poskytovat ucelený pohled na tuto vícevrstvou síť. Aktéry s vysokým stupněm, nebo vysokou hodnotou shlukovacího koeficientu ve flattened síti můžeme považovat za celkově zajímavé. O konkrétních vztazích nám ovšem opět nic neřekne.

### 6.2.6 Zhodnocení experimentů

Jednotlivé vrstvy a flattened síť byly porovnány na celkem 5 vícevrstvých sítích. Z provedených experimentů můžeme jasně vidět, že analýza vícevrstvých sítí pomocí jednotlivých vrstev může být velice zajímavá. Může nám o aktérech a také vrstvách prozradit zajímavé informace. Na druhou stranu při flatteningu většinu těchto informací ztrácíme. Nevíme, co jednotlivé hrany znamenají, a při existenci vrstev, které se navzájem znehodnocují (pozitivní vztahy vůči negativním), se aktéři s vysokými hodnotami měřených vlastností mohou zdát důležití i když ve skutečnosti nejsou. Z těchto důvodů bychom při analýze vícevrstvých sítí měli vždy vědět, co hrany na jednotlivých vrstvách reprezentují a podle toho také k výsledné analýze přistupovat.

## 6.3 Zjištění role aktéra

Cílem tohoto experimentu je zjistit roli vybraného aktéra ve vícevrstvé síti. Aktéři se stejnou nebo podobnou rolí, se spolu často stýkají, provádí podobné aktivity apod. U vybraného aktéra se podíváme na jeho míry na jednotlivých vrstvách a podle těchto výsledků se pokusíme stanovit jeho roli.

### 6.3.1 AUCS

V této síti se můžeme setkat s 61 aktéry, kteří mají 4 různé role. Jedná se o profesory, doktorandy, postdoktorské výzkumníky a administrativní pracovníky. Aktérem, u kterého se pokusíme zjistit jeho roli, bude aktér U4. Nejprve se podíváme, jestli má nějakou vazbu na vrstvě coauthor, což by značilo, že se nejspíše nejedná o administrativního pracovníka. Hodnota neighborhood centrality na vrstvě coauthor u tohoto aktéra je 0. Tento aktér tedy nepublikoval žádný článek, nebo práci. Nejspíše tedy bude administrativní pracovník. Degree centralita tohoto aktéra je 75, je to také zároveň nejvyšší hodnota degree centrality v celé síti. Jedná se tedy o jednoho z nejdůležitějších aktérů z pohledu degree centrality. Jeho neighborhood centralita napříč všemi vrstvami je 26. Maximální hodnota je 60, což je skoro polovina. Téměř polovina aktérů této sítě přišla do kontaktu s tímto aktérem. Nejvyšší hodnotu neighborhood centrality má tento aktér na

vrstvě work. Nejmenší má pak na vrstvě leisure, která reprezentuje volný čas aktérů. Tento aktér tedy nepublikoval žádnou práci nebo článek, je v kontaktu s téměř každým aktérem této sítě, ale netráví s nimi svůj volný čas. Z těchto pozorování můžeme usoudit, že se jedná o jednoho z administrativních pracovníků. Většina aktérů s tímto aktérem přišla do styku, když něco potřebovali zařídit apod. Když se pak podíváme na jeho skutečnou roli, tak skutečně zjistíme, že se jedná o jednoho z 5 administrátorů. Zjistit roli některého z profesorů, postdoktorských výzkumníků, nebo doktorandů by bylo určitě mnohem obtížnější. A to hlavně z toho důvodu, že se zde nevyskytuje žádná vrstva, která by dokázala oddělit tyto role, jak tomu bylo v případě vrstvy coauthor pro administrativní pracovníky.

### 6.3.2 Zhodnocení experimentu

Můžeme si všimnout, že u tohoto experimentu byla použita pouze jedna vícevrstvá síť. Je to z toho důvodu, že tento experiment je velice specifickým, a nelze jej úspěšně provést u většiny vícevrstevných sítí, se kterými jsme zde pracovali. Většina ostatních sítí většinou neměla definované role aktérů, nebo neměla vrstvy, které by umožňovaly oddělit role aktérů. I u sítě AUCS, u které jsme experiment provedli nejsme schopni stanovit roli u každého aktéra. Vrstva coauthor umožňuje oddělit aktéry, kteří publikovali nějakou práci nebo článek a aktéry, kteří se na této vrstvě nevyskytují tedy můžeme považovat za administrativní pracovníky. Následně pomocí analýzy jednotlivých vrstev jsme byli schopni úspěšně stanovit roli zvoleného aktéra. Pokud bychom pro experiment ovšem zvolili aktéra, který se na vrstvě coauthor nachází, nejspíš bychom nebyli schopni rozeznat jestli se jedná o profesora, doktoranda nebo postdokorského výzkumníka.

## 6.4 Porovnání centralit založených na náhodné procházce

V kapitole 6.4 jsme se zmínili o centralitách vícevrstevných sítí, které k výpočtu využívaly náhodnou procházku. Každá z těchto centralit má svůj protějšek i u jednovrstevných sítí, ve kterých náhodnou procházku používat nemusí. Je nutné poznamenat, že centrality jednovrstevných sítí jsou zde počítány na flattened síti. Centrality založené na náhodné procházce jsou počítány na samotné vícevrstvé síti. Každá z porovnávaných centralit bude porovnána pomocí kumulativní doplňkové distribuce a také pomocí tabulky, ve které bude vždy 5 aktérů s nejvyšší hodnotou dané centrality.

### 6.4.1 Florentine Families

V tabulce 14a můžeme vidět 5 aktérů s nejvyšší hodnotou random walk betweenness centrality, v tabulce 14b je 5 aktérů s nejvyšší hodnotou betweenness centrality na flattened síti. Obě tabulky mají celkem 3 společné aktéry, aktér s nejvyšší hodnotou dané centrality je u obou centralit stejný. Na obrázku 24b je grafické znázornění obou verzí těchto centralit pomocí kumulativní doplňkové distribuce. Můžeme si všimnout, že většina aktérů má vyšší hodnoty random walk

betweenness centrality, než betweenness centrality. Toto může být zapříčiněno povahou náhodné procházky, která započítává všechny cesty a ne jen ty nejkratší.

Tabulka 14: Top 5 aktérů vícevrstvé sítě Florentine Families seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality

(a)		(b)	
Actor	rw betweenness	Actor	betweenness
Medici	82	Medici	43,33
Guadagni	62	Barbadori	16,58
Strozzi	58	Guadagni	11,75
Barbadori	56	Ridolfi	7,42
Bischeri	56	Peruzzi	6,42

Pět aktérů s nejvyšší hodnotou random walk closeness můžeme vidět v tabulce 15a, v tabulce 15b je naopak 5 aktérů s nejvyšší hodnotou closeness centrality. V těchto tabulkách tentokrát najdeme pouze jednoho společného aktéra. Grafické znázornění distribucí těchto dvou centralit je na obrázku 24c. Stejně jako v případě předchozích centralit jsou i tyto distribuce velice rozdílné. Aktérů s relativně nízkou hodnotou random walk closeness centrality je zde více, než v případě closeness centrality. Z distribuce closeness centrality můžeme také vidět, že všichni aktéři mají relativně vysokou hodnotu closeness centrality. U random walk closeness centrality se počítá stejně jako u random walk betweenness centrality se všemi cestami. Toto může být opět důvodem relativně nižších hodnot random walk closeness centrality u některých aktérů.

Tabulka 15: Top 5 aktérů vícevrstvé sítě Florentine Families seřazených podle (a) random walk closeness centrality, (b) closeness centrality

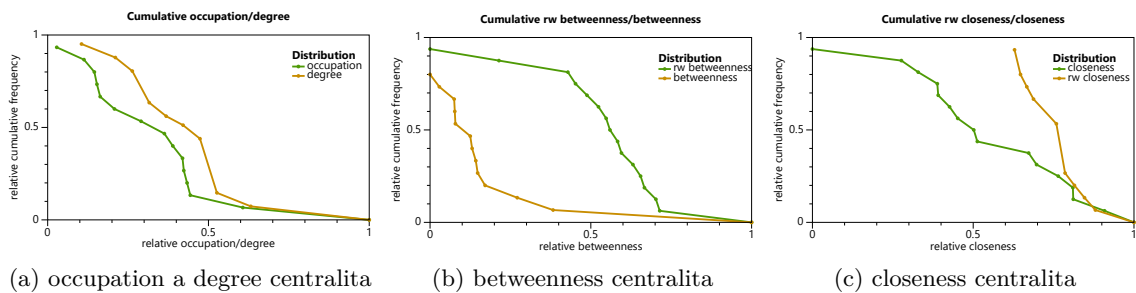
(a)		(b)	
Actor	rw closeness	Actor	closeness
Bischeri	0,29	Medici	0,64
Guadagni	0,29	Barbadori	0,56
Medici	0,28	Ridolfi	0,54
Lamberteschi	0,21	Tornabuoni	0,52
Pazzi	56	Ginori	0,50

Poslední z porovnávaných dvojic centralit je occupation a degree. Tabulka 16a obsahuje 5 aktérů s nejvyšší hodnotou occupation centrality, tabulka 16b obsahuje pět aktérů s nejvyšší hodnotou degree centrality. Obě tabulky mají celkem 4 společné aktéry. Aktér s nejvyšší hodnotou

dané centrality je v obou tabulkách stejný, stejně jako v případě centralit random walk betweenness centrality a betweenness centrality. Na obrázku 24a je grafické znázornění kumulativních doplňkových distribucí těchto dvou centralit. Na rozdíl od předchozích dvou párů porovnávaných centralit jsou si distribuce těchto dvou centralit velice podobné.

Tabulka 16: Top 5 aktérů vícevrstvé sítě Florentine Families seřazených podle (a) occupation centrality, (b) degree centrality

(a)		(b)	
Actor	occupation	Actor	degree
Medici	483	Medici	19
Guadagni	262	Peruzzi	12
Peruzzi	234	Barbadori	10
Barbadori	220	Bischeri	10
Castellani	213	Castellani	10



Obrázek 24: Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě Florentine Families

#### 6.4.2 Sampson's Monastery

Tabulky 17a a 17b obsahují pět aktérů s nejvyšší hodnotou random walk betweenness centrality a betweenness centrality. Aktér s nejvyššími hodnotami daných centralit se tentokrát neshoduje, liší se o jednu pozici. Tabulky obsahují celkem 2 společné aktéry, jistá podobnost mezi výsledky těchto centralit se zde opět objevuje. Na obrázku 25b je grafické znázornění distribucí těchto dvou centralit. Můžeme si všimnout, že stejně jako u sítě Florentine Families, je zde relativně více aktérů s vyšší hodnotou random walk betweenness centrality.

Aktéry s 5 nejvyššími hodnotami random walk closeness centrality a closeness centrality můžeme vidět v tabulkách 18a a 18b.

Tabulka 17: Top 5 aktérů vícevrstvé sítě Sampsons's Monastery seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality

(a)		(b)	
Actor	rw betweenness	Actor	betweenness
Peter_4	832	Victor_8	1,71
Bonaven_5	764	Peter_4	1,50
Romul_10	761	Basil_3	1,44
Greg_2	757	Simp_18	1,36
Basil_3	738	Louis_11	1,34

Tabulka 18: Top 5 aktérů vícevrstvé sítě Sampsons's Monastery seřazených podle (a) random walk closeness centrality, (b) closeness centrality

(a)		(b)	
Actor	rw closeness	Actor	closeness
Bonaven_5	0,26	Victor_8	0,94
Boni_15	0,24	Boni_15	0,85
Ambrose_9	0,23	John_1	0,85
Greg_2	0,20	Louis_11	0,85
Mark_7	0,19	Amand_13	0,85

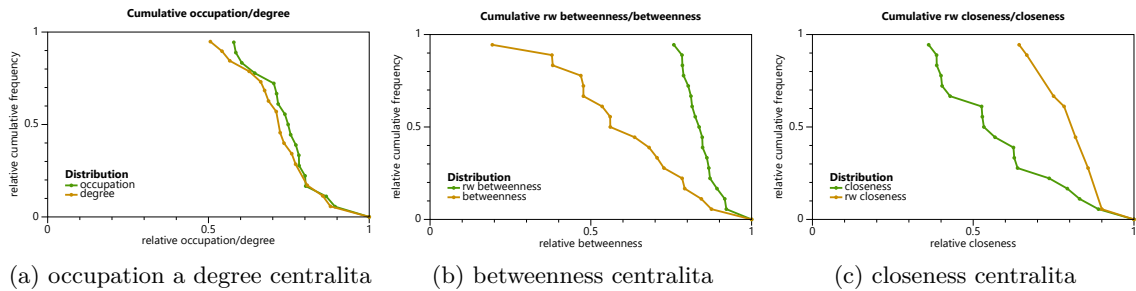
Tabulka 19: Top 5 aktérů vícevrstvé sítě Sampsons's Monastery seřazených podle (a) occupation centrality, (b) degree centrality

(a)		(b)	
Actor	occupation	Actor	degree
Peter_4	2985	Peter_4	83
Greg_2	2805	Greg_2	73
Elias_17	2487	Basil_3	71
Basil_3	2451	John_1	67
Simp_18	2336	Simp_18	67

V obou tabulkách nalezneme pouze jednoho společného aktéra, shodou okolností na stejné pozici. Při pohledu na distribuce těchto dvou centralit, můžeme vidět, podobné rozdíly jako u sítě

Florentine Families. Opět je zde více aktérů s relativně nižší hodnotou random walk closeness centrality a většina aktérů má relativně vysokou closeness centralitu.

V tabulkách 19a a 19b můžeme spatřit 5 aktérů s nejvyššími hodnotami occupation centrality a degree centrality. Tabulky mají celkem 4 společné aktéry, aktér s nejvyšší i druhou nejvyšší hodnotou dané centrality je zde stejný. Distribuce těchto dvou centralit jsou téměř totožné (viz obrázek 25a). Podobnost těchto dvou centralit je opět velmi vysoká.



Obrázek 25: Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě Sampson's Monastery

### 6.4.3 Bank Wiring

Tabulky 20a a 20b obsahují 5 aktérů s nejvyšší hodnotou random walk betweenness centrality a betweenness centrality. Tabulky obsahují celkem 2 společné aktéry, aktér s nejvyšší ani druhou nejvyšší hodnotou dané centrality se zde ovšem neshoduje. Distribuce těchto dvou centralit jsou velmi podobné předchozím vícevrstvým sítím (viz obrázek 26b).

Při porovnání random walk closeness centrality a closeness centrality si můžeme všimnout, že obě tabulky mají celkem 2 společné aktéry. Aktér s nejvyšší hodnotou random walk closeness centrality je zde až druhým aktérem s nejvyšší hodnotou closeness centrality (viz tabulky 21a a 21b). Při porovnání distribucí těchto centralit můžeme vidět velice podobnou situaci jako u předchozích vícevrstvých sítí (viz obrázek 26c).

Tabulka 20: Top 5 aktérů vícevrstvé sítě Bank Wiring seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality

(a)		(b)	
Actor	rw betweenness	Actor	betweenness
I3	211	W5	9,26
W3	199	W6	6,06
S4	175	W4	3,76
W2	171	W2	3,32
W4	170	S1	2,74

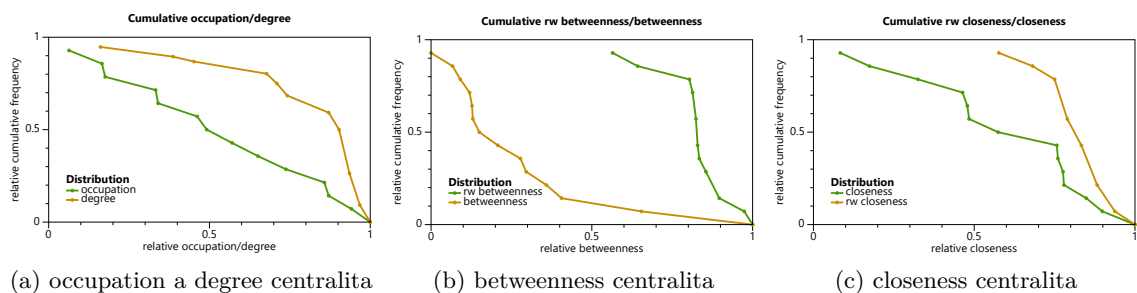
Pět aktérů s nejvyššími hodnotami occupation centrality a degree centrality můžeme vidět v tabulkách 22a a 22b. Tabulky mají celkem 3 společné aktéry. Při pohledu na distribuce těchto dvou centralit (viz obrázek 26a) můžeme vidět, že se tyto distribuce tentokrát relativně odlišují. Aktérů s relativně nižší occupation centralitou je zde relativně více, než naopak s degree centralitou.

Tabulka 21: Top 5 aktérů vícevrstvé sítě Bank Wiring seřazených podle (a) random walk closeness centrality, (b) closeness centrality

(a)		(b)	
Actor	rw closeness	Actor	closeness
S1	0,24	W5	0,87
W2	0,21	S1	0,81
W4	0,20	W60	0,81
W7	0,20	W7	0,76
W8	0,20	W9	0,76

Tabulka 22: Top 5 aktérů vícevrstvé sítě Bank Wiring seřazených podle (a) occupation centrality, (b) degree centrality

(a)		(b)	
Actor	occupation	Actor	degree
S4	1039	W7	31
S1	915	S1	30
W7	838	W6	30
W8	782	W5	29
W5	625	W9	29



Obrázek 26: Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě Bank Wiring

#### 6.4.4 Tailor Shop

Random walk betweenness a betweenness centrality můžeme opět porovnat pomocí tabulek 23a a 23b. V tabulkách můžeme vidět pět aktérů s nejvyššími hodnotami dané centrality. Pokud se na obě tabulky podíváme, můžeme si všimnout, že aktér s nejvyšší hodnotou dané centrality se zde neshoduje. Obě tabulky ovšem mají celkem 4 společné aktéry, což značí relativně vysokou podobnost. Při pohledu na distribuce těchto dvou centralit můžeme vidět opět velice podobnou situaci jako u všech předchozích sítí (viz 27b). Počet aktérů s relativně vysokou hodnotou random walk betweenness je zde opět vyšší, než v případě betweenness centrality.

Tabulky 24a a 24b obsahují pět aktérů s nejvyššími hodnotami random walk closeness centrality a closeness centrality. V obou tabulkách nalezneme pouze jednoho společného aktéra. Na obrázku 27c můžeme vidět distribuce těchto dvou centralit. Stejně jako v předchozích případech je i zde closeness centralita relativně vyšší pro většinu aktérů.

Occupation centralitu a degree centralitu můžeme porovnat pomocí tabulek 25a a 25. Obě tabulky mají všech pět aktérů společných, liší se pouze v pořadí prvních dvou aktérů. Při pohledu na distribuce těchto dvou centralit můžeme opět vidět výraznou podobnost (viz obrázek 27a).

Tabulka 23: Top 5 aktérů vícevrstvé sítě Tailor Shop seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality

(a)		(b)	
Actor	rw betweenness	Actor	betweenness
Lyashi	1815	Mukubwa	57,57
Chisokone	1770	Lyashi	48,10
Mukubwa	1672	Chisokone	42,19
Abraham	1582	Hastings	27,83
Ibrahim	1576	Ibrahim	16,02

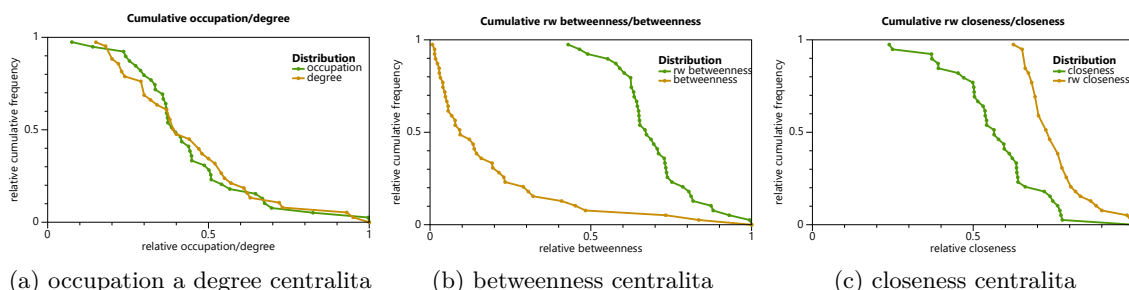
Tabulka 24: Top 5 aktérů vícevrstvé sítě Tailor Shop seřazených podle (a) random walk closeness centrality, (b) closeness centrality

(a)		(b)	
Actor	occupation	Actor	degree
Mabange	0,12	Mukubwa	0,84
Mpundu	0,12	Lyashi	0,84
Hastings	0,12	Chisokone	0,83
Nkoloya	0,11	Hastings	0,76
Angel	0,11	Meshak	0,75



Tabulka 25: Top 5 aktérů vícevrstvé sítě Tailor Shop seřazených podle (a) occupation centrality, (b) degree centrality

(a)		(b)	
Actor	rw closeness	Actor	closeness
Lyashi	2410	Chisokone	100
Chisokone	2211	Lyashi	95
Mukubwa	2119	Mukubwa	93
Abraham	1815	Abraham	73
Ibrahim	1642	Ibrahim	72



Obrázek 27: Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě Tailor Shop

#### 6.4.5 AUCS

Tabulky 26a a 26b opět obsahují pět aktérů s nejvyššími hodnotami random walk betweenness centrality a betweenness centrality. Pokud se na tabulky podíváme pozorněji, můžeme si všimnout, že obě tabulky obsahují celkem 3 společné aktéry. Jejich pořadí je ovšem odlišné. Při pohledu na distribuce těchto dvou centralit (viz obrázek 28b) si můžeme všimnout podobné situace jako u všech předchozích vícevrstvých sítí.

Random walk closeness centralitu a closeness centralitu můžeme porovnat pomocí tabulek 27a a 27b. Kde každá tabulka opět obsahuje pět aktérů s nejvyšší hodnotou příslušné centrality. Obě tabulky tentokrát obsahují pouze jednoho společného aktéra. Na obrázku 28c můžeme vidět distribuce těchto dvou centralit. Situace se ani zde výrazně nezměnila. Počet aktérů s relativně vysokou closeness centralitou je opět vyšší, než v případě random walk closeness centrality.

Poslední z porovnávaných dvojic centralit je occupation a degree. Tabulka 28a obsahuje 5 aktérů s nejvyšší hodnotou occupation centrality, tabulka 28b obsahuje pět aktérů s nejvyšší hodnotou degree centrality. Tabulky obsahují celkem 3 společné aktéry, kteří se liší v pořadí. Při pohledu na obrázek 28a si můžeme všimnout, že si jsou tyto dvě distribuce opět, jak už tomu bylo u několika předchozích vícevrstvých sítí, velice podobné.

Tabulka 26: Top 5 aktérů vícevrstvé sítě AUCS seřazených podle (a) random walk betweenness centrality, (b) betweenness centrality

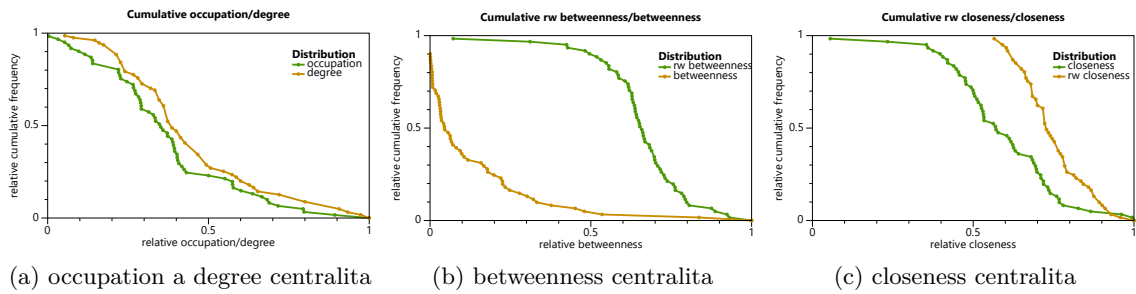
(a)		(b)	
Actor	rw betweenness	Actor	betweenness
U123	3951	U4	235,34
U67	3678	U123	196,90
U4	3677	U91	125,59
U126	3532	U67	112,95
U130	3495	U79	106,25

Tabulka 27: Top 5 aktérů vícevrstvé sítě AUCS seřazených podle (a) random walk closeness centrality, (b) closeness centrality

(a)		(b)	
Actor	rw closeness	Actor	closeness
U110	0,10	U123	0,66
U3	0,09	U4	0,63
U91	0,09	U79	0,61
U67	0,09	U67	0,61
U54	0,09	U130	0,60

Tabulka 28: Top 5 aktérů vícevrstvé sítě AUCS seřazených podle (a) occupation centrality, (b) degree centrality

(a)		(b)	
Actor	occupation	Actor	degree
U91	1565	U4	75
U110	1552	U123	73
U1	1366	U67	70
U4	1345	U79	68
U67	1285	U91	68



Obrázek 28: Kumulativní relativní doplňkové distribuce occupation, degree, betweenness a closeness centralit vícevrstvé sítě AUCS

#### 6.4.6 Zhodnocení experimentů

V těchto experimentech bylo porovnáno celkem 5 vícevrstevných sítí. U každé sítě byly porovnávány dvojice centralit random walk betweenness a betweenness, random walk closeness a closeness a occupation a degree. U první z porovnávaných párů centralit jsme si mohli všimnout, že hodnoty aktérů u této centrality byly relativně vyrovnanější v porovnání s betweenness centralitou. U random walk betweenness centrality jsme se téměř nikdy nesetkali s aktéry, kteří by měli relativně nízkou hodnotu oproti aktérům s nejvyššími hodnotami. U betweenness centrality bylo ovšem běžné, že někteří aktéři měli výrazně vyšší hodnoty, než ostatní. Toto může být způsobeno hlavně povahou náhodné procházky, která zohledňuje i aktéry, přes které neprochází téměř žádné nejkratší cesty. U druhého páru porovnávaných centralit jsme viděli velice podobný výsledek. Většina aktérů měla vždy relativně nižší hodnotu random walk closeness centrality, než naopak closeness centrality. Tyto výsledky můžeme opět považovat za důsledek náhodné procházky, která počítá se všemi cestami a ne pouze s těmi nejkratšími. Aktérů s relativně vysokou hodnotou random walk closeness centrality je pak tedy méně. Poslední z porovnávaných párů měl mezi sebou největší podobnost. Téměř u každé vícevrstvé sítě, u které jsme porovnávali occupation a degree centralitu, mělo více než 3 stejné aktéry v tabulkách, které zachycovaly 5 aktérů s nejvyššími hodnotami daných centralit. Distribuce těchto dvou centralit byly také často velmi podobné, téměř až identické u některých vícevrstevných sítí. Porovnávané páry centralit si tedy jsou do jisté míry podobné, při výsledné analýze ale musíme brát v potaz výše zmíněné odlišnosti a podle toho výslednou centralitu zvolit.

## 7 Závěr

Analýza vícevrstvých sociálních sítí je novou disciplínou, se kterou se v současnosti při analýze sociálních sítí často nesetkáme. Cílem této práce bylo se s těmito sítěmi a jejich analýzou seznámit, implementovat nástroj, který by umožňoval jejich analýzu pomocí několika možných přístupů a následně nad těmito sítěmi provést experimenty. Cílem experimentů bylo porovnat jednotlivé přístupy analýzy vícevrstvých sociálních sítí a zjistit, který z přístupů je kdy vhodné použít. V rámci práce byly nejprve popsány jednovrstvé sítě a jejich vlastnosti, následně byly popsány samotné vícevrstvé sítě včetně možných přístupů při jejich analýze a jejich vlastností. Výsledný implementovaný nástroj umožňuje analýzu vícevrstvých sítí pomocí několika přístupů. Byly implementovány jak centrality, které se běžně využívají u jednovrstvých sítí, tak centrality, se kterými se setkáme pouze u vícevrstvých sítí. Nástroj umožňuje analýzu samotné vícevrstvé sítě a jejich globálních vlastností, tak analýzu jednotlivých vrstev. Tyto vrstvy je pak možné mezi sebou porovnávat pomocí distribucí jejich vlastností.

Stávající nástroj určitě poskytuje prostor pro zlepšení. Nástroj byl implementován v jazyce C#, tento jazyk, v porovnání s jinými jazyky, neposkytuje takové množství knihoven pro analýzu sítí a jejich vizualizaci. Mezi budoucí rozšíření nástroje bychom mohli zahrnout další algoritmy a varianty stávajících algoritmů pro výpočet vlastností jak jednovrstvých, tak vícevrstvých sítí. Nástroj také není přizpůsoben pro práci s orientovanými ohodnocenými sítěmi, a také nejsou implementovány algoritmy pro práci s komunitami, zde se tedy také nabízí možné budoucí rozšíření.

## Literatura

- [1] DICKISON, Mark E., Matteo MAGNANI a Luca ROSSI. *Multilayer social networks*. New York, NY: Cambridge University Press, 2016. ISBN 978-1107438750.
- [2] NEWMAN, Mark. *Networks: An Introduction*. New York: Oxford University Press, 2010. ISBN 01-992-0665-1
- [3] BARABÁSI, Albert-László. a Márton. PÓSFAL. *Network science*. Cambridge, United Kingdom: Cambridge University Press, 2016. ISBN 978-110-7076-266
- [4] Leonard Euler's Solution to the Königsberg Bridge Problem. PAOLETTI, Teo. *Mathematical Association of America* [online]. [cit. 2018-03- 25].
- [5] KOVÁŘ, Petr. *Úvod do teorie grafů* [online]. [cit. 2018-03-25]. Dostupné z: [http://home1.vsb.cz/~kov16/files/uvod\\_do\\_teorie\\_grafu.pdf](http://home1.vsb.cz/~kov16/files/uvod_do_teorie_grafu.pdf)
- [6] ČERNÝ, Jakub. *Základní grafové algoritmy* [online]. KAM, MFF UK, 2010 [cit. 2018-03-25]. Dostupné z: <http://kam.mff.cuni.cz/~kuba/ka/ka.pdf>
- [7] SLANINA, František a Miroslav KOTRLA. *Sítě „malého světa“: Proč mají odlišné sítě podobnou strukturu? Vesmír*. 2001, 2001(11), 611-614. Dostupné z: <http://casopis.vesmír.cz/files/file/fid/1198/aid/4791>
- [8] BARABÁSI, Albert-László. *Linked: how everything is connected to everything else and what it means for business, science, and everyday life*. New York: Basic Books, 2014. ISBN 04-650-8573-3.
- [9] BHAGAT, Smriti, Moira BURKE, Carlos DIUK, Ismail ONUR FILIZ a Sergey EDUNOV. *Three and a half degrees of separation*. In: *Facebook Research* [online]. 2016 [cit. 2018-03-25]. Dostupné z: <https://research.fb.com/three-and-a-half-degrees-of-separation/>
- [10] HOLME, Petter, Jari SARAMÄKI. *Temporal networks*. New York: Springer, 2013. ISBN 978-3642364600.
- [11] GAMMA, Erich. *Design patterns elements of reusable object-oriented software*. Reading: Addison-Wesley, c1995. Addison-Wesley professional computing series. ISBN 0-201-63361-2.
- [12] *The MIT License / Open Source Initiative*. [online]. [cit. 2018-03-30]. Dostupné z: <https://opensource.org/licenses/MIT>
- [13] SYROMIATNIKOV, Artem a Danny WEYNS. *A Journey Through the Land of Model-View-\* Design Patterns* [online]. , 10 [cit. 2018-04-03]. Dostupné z: <http://homepage.lnu.se/staff/daweaa/papers/2014WICSA.pdf>

- [14] ERDŐS, Paul a Alfréd RÉNYI. *On random graphs*. I. Publicationes Mathematicae 6, 1959, 290-297.
- [15] TAHA, Samah M. Osman. *Small-World Network Models and Their Average Path Length* [online]. [cit. 2018-04-08]. Dostupné z: [https://scholar.sun.ac.za/bitstream/handle/10019.1/95834/taha\\_smallworld\\_2014.pdf](https://scholar.sun.ac.za/bitstream/handle/10019.1/95834/taha_smallworld_2014.pdf)
- [16] *Wikipedia Articles live in a Small World* [online]. [cit. 2018-04-08]. Dostupné z: <https://blogs.cornell.edu/info2040/2011/09/27/1598/>
- [17] BARABÁSI, Albert-László. *V pavučině sítí*. V Praze: Paseka, 2005. Fénix (Paseka). ISBN 80-7185-751-3.
- [18] LOVÁSZ, László. *Random Walks on Graphs: A Survey* [online]. [cit. 2018-04-11]. Dostupné z: <http://www.leonidzhukov.net/hse/2017/networkscience/papers/LovaszRadnomWalks93.pdf>
- [19] LUCA, Rossi a Magnani MATTEO. *Towards effective visual analytics on multiplex and multilayer networks: Chaos, Solitons and Fractals.*, 2015. 68-76.

## A Příloha na CD/DVD

- **src** - zdrojové kódy implementace
- **setup** - soubory pro instalaci aplikace
- **dp.pdf** - text této práce
- **data** - použité datové sady